REGULATION OF V(D)J RECOMBINATION: MECHANISMS THAT GOVERN LOCUS ACCESSIBILITY AND MITIGATE AGAINST GENOMIC INSTABILITY

by
Meiling Rose-Anne May

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
July 2020

Abstract

V(D)J recombination is the process that generates the primary antigen receptor repertoire of the adaptive immune system. This process involves the ordered rearrangement of gene segments within antigen receptor loci and is initiated by the V(D)J recombinase, comprised of the recombination activating gene 1 and 2 proteins (RAG-1 and RAG-2). V(D)J recombination involves the programmed introduction of DNA double stranded breaks in the genome and is therefore tightly regulated to ensure genomic stability. Epigenetic regulation of V(D)J recombination by trimethylation at lysine 4 of histone 3(H3K4me3) is modulated through the plant homeodomain (PHD) of RAG-2. Binding of the RAG-2 PHD to H3K4me3 couples V(D)J recombination to active transcription and H3K4me3 allosterically activates RAG. In this dissertation, we demonstrate that the binding and allosteric activation of H3K4me3 are separable functions of the RAG-2 PHD.

Temporal regulation of V(D)J recombination is an additional method of preserving genomic integrity. RAG-2 undergoes periodic destruction at the G1-to-S transition through phosphorylation at T490 and subsequent degradation by the ubiquitin-proteasome pathway. This limits RAG activity to the G1 cell cycle phase, during which nonhomologous end joining (NHEJ) is the predominant form of DNA double stranded break repair. In this dissertation, we employ a mouse model in which RAG-2 protein is constitutively expressed and DNA double stranded breaks accumulate throughout the cell cycle. To characterize genomic rearrangements that result from misregulated RAG activity, we developed a capture-sequencing approach. We demonstrate that this method is capable of identifying endogenous V(D)J rearrangements and novel translocations involving antigen receptor loci.

Ph.D. Dissertation referees for Meiling Rose-Anne May

Stephen Desiderio, M.D., Ph.D.

Director Emeritus, Institute for Basic Biomedical Sciences

Professor (Emeritus), Molecular Biology and Genetics

Professor (Emeritus), Medicine

Johns Hopkins University School of Medicine

(Thesis advisor)

Carol Greider, Ph.D.

Chair of department, Molecular Biology and Genetics

Bloomberg Distinguished Professor, Molecular Biology and Genetics

Johns Hopkins School of Medicine

(Thesis co-advisor)

Ranjan Sen, Ph.D.

Professor (Adjunct), Medicine

Johns Hopkins University School of Medicine

Chief, Laboratory of Cellular and Molecular Biology

National Institute on Aging

(reader)

Acknowledgements

Dr. Stephen Desiderio has been my scientific role model and I aspire to be like him. I chose his lab because I wanted to learn to think about science, design controlled experiments, and perform rigorous science like Steve. I looked to the students that went through his lab and saw how differently and rigorously they approached science compared to students in other laboratories. I knew that this was a direct result of training under Steve. His encouragement, commitment, and guidance throughout my dissertation helped me to accomplish this milestone and gave me the independence to pursue scientific ideas and professional development opportunities. I could not have asked for a more supportive mentor regarding my professional development as Steve let me do a teaching fellowship and an internship at AstraZeneca. He believed in me and my abilities and trusted that even with these time-consuming professional development opportunities that I would be committed to my project and get my work done.

Dr. Carol Greider so kindly accepted me into her laboratory as a rotation student and then later as a guest graduate student. She was welcoming and treated me just like one of her own lab members even though I was working on V(D)J recombination and not telomeres. She has always supported me and has been a huge proponent for women in science. I strive to be a leader like Carol, who is a great scientist yet humble and approachable.

My dissertation experience would not have been the same if it weren't for my Desiderio lab mates (John Bettridge, Nick Dordai, Wenyan Lu, and Alyssa Ward) and my Greider lab mates (Kayarash Karimian, Margaret Strong, Callie Shubin, Sam Sholes, Rini Mayangsari, Carla Connelly, Christine Gao). I cannot thank you enough for the endless scientific discussions,

commiseration sessions over failed experiments, and countless laughs. From game nights in the conference room to walks to the hospital to look at poster proofs, I've enjoyed every moment spent with you. I've learned so much from you scientifically, like protein preparations, southern blots, and scientific resilience, but more importantly I've learned how to be a good friend and colleague, how to support each other during the failed experiments and during the triumphs of positive data.

I'd like to thank my thesis committee Carol Greider, Kathy Burns, Ranjan Sen, and Sarah Wheelan for their unwavering support of both my science and my professional aspirations. I always left my committee meetings with much clearer direction and wishing I had met with you earlier. You volunteered so much support and offered technical assistance from your lab members.

I could not have finished this dissertation without the help and support of the MBG department and the BCMB graduate program. MBG has been such a collaborative environment and through journal clubs and colloquia have helped me to understand topics outside my field and to hone my publication skills. Jess Terzigni has been instrumental in easing my transition between laboratories, in and out of internship, and through the publication process. BCMB, particularly Arhonda Gogos and Sharon Root have been a constant source of support for me throughout all the ups and downs of graduate school. My BCMB classmates are amazing scientists and I've learned so much from them both inside and outside of the laboratory.

Lastly, I'd like to acknowledge my friends and family (especially Kevin) for being a constant source of support during this journey. Despite not always understanding the science, you have always been there for me to boost my confidence and believe that I can accomplish great things.

Table of Contents

List of Tables

List of Figures

Chapter 1: Binding and allosteric transmission of histone 3 Lys-4 trimethylation to the recombinase RAG-1 are separable functions of the RAG-2 PHD finger

Abstract

V(D)J recombination is initiated by the recombination-activating gene (RAG) recombinase, consisting of RAG-1 and RAG-2 subunits. The susceptibility of gene segments to cleavage by RAG is associated with gene transcription and with epigenetic marks characteristic of active chromatin, including trimethylation of histone H3 at lysine 4 (H3K4me3). Binding of H3K4me3 by a plant homeodomain (PHD) in RAG-2 induces conformational changes in RAG-1, allosterically stimulating substrate binding and catalysis. To better understand the path of allostery from the RAG-2 PHD finger to RAG-1, here we employed phylogenetic substitution. We observed that a chimeric RAG-2 protein in which the mouse PHD finger is replaced by the corresponding domain from the shark *Chiloscyllium punctatum* binds H3K4me3, but fails to transmit an allosteric signal, indicating that binding of H3K4me3 by RAG-2 is insufficient to support recombination. By substituting residues in the *C. punctatum* PHD with the corresponding residues in the mouse and testing for rescue of allostery, we demonstrate that H3K4me3 binding and transmission of an allosteric signal to RAG-1 are separable functions of the RAG-2 PHD finger.

Introduction

The human immune system

The human body is continually challenged by viruses, bacteria, and other pathogens. To prevent and eliminate infections, the immune system has evolved to recognize, clear, and remember pathogens. The human immune system is comprised of two arms, a nonspecific but rapid response called the innate immune system and a highly specific but slower response called the adaptive immune system. Innate immunity consists of both physical barriers like skin and mucus, as well as cellular and protein systems such as white blood cells, the complement system, and nucleic acid sensors that recognize pathogen associated molecular patterns. The highly antigen specific response of the adaptive immune system is carried out by lymphocytes, specifically B cells and T cells. B cells are capable of producing antigen-specific antibodies and B cell receptors (BCRs), collectively termed immunoglobulins (Ig), whereas T cells produce antigen-specific T cell receptors (TCRs) that recognize foreign peptides. Ig and TCR contain variable domains that are responsible for binding antigen, and constant domains responsible for effector functions. The genes that encode Ig and TCR are assembled by the process of V(D)J recombination. Igs are comprised of light and heavy chain polypeptides encoded by the immunoglobulin heavy chain locus (IGH) and the immunoglobulin □ or □ light chain loci (IGK, IGL) respectively. TCRs are encoded by the TRB, TRAD, and TRG loci. These loci specify polypeptides that assemble into αβ or γδ heterodimers.

Overview of V(D)J recombination

Antigen receptor genes are present in the germline as discrete segments, termed variable (V),

diversity (D), and joining (J), that are joined during lymphocyte development by V(D)J recombination (1). V(D)J recombination is initiated by a specialized transposase (2), composed of RAG-1 and RAG-2 subunits, which cleaves DNA at recombination signal sequences (RSSs) that flank the participating gene segments (3). There are two classes of RSSs, termed 12-RSS and 23-RSS, in which heptamer and nonamer elements are separated by spacers of 12 bp or 23 bp. DNA cleavage requires the pairing of a 12-RSS with a 23-RSS (3). Rearrangement is initiated by nicking of DNA by RAG at the junction of each gene segment with its flanking RSS, followed by transesterification to produce double-strand breaks. The resulting coding and signal ends are then joined by non-homologous end joining to produce coding joints and signal joints (1,4). The catalytic and DNA-binding functions reside largely within RAG-1, but RAG-2 is also essential for DNA cleavage, contributing to DNA-binding and stabilization of the active site (5). The carboxy-terminal quarter of RAG-2 is dispensable for DNA cleavage but carries out several regulatory functions, including binding of RAG-2 to H3K4me3, an epigenetic mark of active chromatin (6-8). V(D)J recombination is preceded by the appearance of non-coding transcripts at the unrearranged locus (9,10) and the deposition of H3K4me3 (6,7,11-15), which binds to a plant homeodomain (PHD) finger spanning residues 415 through 487 of mouse RAG-2 (8,16). Engagement of H3K4me3 promotes recombination in vivo (6,7) and stimulates coupled cleavage of RSS substrates in vitro (17-19), consistent with allosteric activation. In agreement with this interpretation, binding of H3K4me3 by the RAG-2 PHD finger is associated with changes in the conformational distribution of RAG-1 (20).

The V(D)J recombinase: RAG-1 and RAG-2

Initial DNA transfer experiments suggested that the V(D)J recombinase was encoded by a single

gene expressed in lymphoid cells (21). Through a series cDNA transfection experiments, the gene encoding RAG-1 was identified and shown to be critical for V(D)J recombination (22). The RAG-1 protein contains 1040 amino acid residues in mouse and 1043 amino acid residues in human (22). Mouse and human RAG-1 coding regions share 86% identity and close to 95% similarity indicating a high degree of evolutionary conservation. Similar genes were identified soon thereafter in rabbit, dog, goat, and horse (22).

Upon further inspection, the coding sequence of RAG-1 was found to be the product of a single exon that was much smaller than the active, 12 kb genomic clone that was initially described. This, together with evidence that the RAG-1 cDNA was no more efficient in activation of recombination than bulk genomic DNA, suggested that the active genomic clone contained a second, tightly linked gene that was required in addition to RAG-1 for V(D)J recombination. A cDNA clone derived from this second gene, encoding a protein termed RAG-2, was found to act synergistically with RAG-1 cDNA to activate V(D)J recombination (23). The genes encoding RAG-1 and 2 are convergently transcribed and only 8 kb apart on chromosome 11p of humans and chromosome 2p of mice (24). The method used to identify the V(D)J recombinase would not have been fruitful if the two genes were not located together in the genome.

Both RAG-1 and RAG-2 are essential for V(D)J recombination. RAG-2 deficient mice exhibit severe combined immunodeficiency (SCID), characterized by complete loss of mature B or T lymphocytes and a lack of antigen receptor loci rearrangements (25). This deficiency could be rescued in cell lines by the introduction of wild-type RAG-2. Since the only abnormal phenotype observed in these mice was the SCID phenotype, this data suggested that RAG-2 function and

V(D)J recombination are not required for development of cells outside of the lymphoid lineage (25). RAG-1 deficient mice also exhibit SCID, have no mature B or T cells, and exhibit a lack of antigen receptor rearrangements (26). Together these RAG deficient mouse models demonstrate the importance of RAG-1 and RAG-2 in V(D)J recombination.

Evolution of RAG-1 and RAG-2

Mechanistically V(D)J recombination is very similar to transposition. RAG-1 and RAG-2 may have evolved from components of a transposable element about 450 million years ago (27). This notion is supported by the observation that in vitro RAG-1 and RAG-2 are capable of acting as a transposase, excising a piece of DNA containing RSS from a donor site and inserting the DNA into a target DNA molecule (28). This results in a short duplication of the target DNA just as transposition events are characterized by target site duplications. Although RAG-1 and RAG-2 can act as a transposase in vitro, there has not been much evidence of this activity in vivo. This could be explained by the additional levels of regulation imposed on this system in vivo.

Several regulatory mechanisms may suppress transposition in vivo. One mode of regulation is GTP-mediated inhibition. RAG-mediated transposition is suppressed by the physiological concentration (0.5 mM) of guanine nucleotide (GTP), which blocks the capture of target DNA (29). RAG transposition in vitro is inhibited by GTP at concentrations of 1mM or higher, which blocks target capture in the signal-end complex(30). This mode of regulation is also found for other transposases such as Tn7, bacteriophage Mu, and Drosophila P element(30). While one study has suggested that the non-core region of RAG-2 can suppress transposition (30), this has been contradicted by a subsequent study in which core and full-length RAG signal end complexes exhibited similar transposition frequencies (31). Magnesium at physiologic

concentrations of 20 - 25mM can suppress transposition by favoring the disintegration of branched transposition intermediates, essentially reversing the process of transposition(30). A recent structure of RAG in a DNA-strand transfer complex revealed the presence of sharp kinks and twists in the DNA which could attenuate transposition(32).

The identification of putative ancestral RAG proteins further supports the transposon origins of RAG. Two ancestral RAG-1 proteins, Transib transposase and purple sea urchin RAG-1 like, when coexpressed with RAG-2 are capable of initiating V(D)J recombination in vitro(33). Additionally, ProtoRAG, a DNA transposon from lancelets, is capable of mediating transposon excision, host DNA recombination, and transposition utilizing mechanisms similar to those employed by vertebrate RAG(34). These findings are consistent with a transposon origin of RAG; however, it remains possible that RAG-1 and RAG-2 may have independent evolutionary origins. RAG-1 is capable of mediating low levels of recombination in the absence of RAG-2, which permits the possibility that the two proteins may have evolved separately(35).

Domain Structure of RAG-1 and RAG-2

Murine RAG-1 and RAG-2 contain 1008 and 527 amino acid residues, respectively, of which residues 384-1008 of RAG-1 and 1-352 of RAG-2 are termed the core regions (36). These are the minimal regions required to perform RSS-dependent DNA cleavage in vitro(37,38). The core of RAG-1 is comprised of the nonamer binding domain (NBD), dimerization and DNA binding domain (DDBD), pre-RNase H fold (PreR) domain, the catalytic RNase H fold (RNH), two zinc binding domains ZnC2 and ZnH2, and a C terminal domain (CTD)(5). The core of RAG-1 contains key catalytic residues needed for catalysis, a triplet of acidic residues commonly called

the DDE motif, which coordinate the divalent metal ion needed for catalysis(39-41). RAG-1 specifically recognizes and binds to the recombination signal sequences (RSSs) in DNA(42-44). This recognition is mediated primarily through the NBD and DDBD domains within core RAG-1(5). While RAG-2 is required for high-affinity binding to RSSs, recent structural data indicate that RAG-1 is largely if not solely responsible for binding DNA(5,45). The core of RAG-2 is comprised of six Kelch like motifs(5). In addition to these core domains, both proteins have extensive domain structures not included within the core domains.

Beyond the core, RAG-1 contains several domains important for protein localization and regulation of V(D)J recombination. The N terminal region can enhance V(D)J recombination (46). There are several domains responsible for nuclear localization of RAG-1. Some of these regions serve as binding sites for the SRP1/Rch1 family of proteins. Rch1 specifically interacts with RAG-1 and is 47% identical to yeast SRP1, which is associated with the nuclear envelope(47). Once in the nucleus, RAG-1 has been found to periodically localize to the nucleolus mediated through RNA interactions with RAG-1 domains containing basic amino acids(48). This nucleolar sequestration may serve to suppress V(D)J recombination(49). RAG-1 contains a RING finger domain that acts as a ubiquitin ligase and can mediate auto-ubiquitylation at a conserved Lys residue, K233 (50); disruption of this RING finger domain is associated with immunodeficiency(50). Furthermore, auto-ubiquitination of RAG-1 at residue K233 is associated with increased DNA cleavage activity(51). Together, this evidence suggests that the non-core regions of RAG-1 are important for both proper RAG-1 function and regulation of V(D)J recombination.

The non-core domains of RAG-2 also play important roles in the mechanism and regulation of V(D)J recombination. Early mutational studies revealed that the non-core carboxy-terminal portion of RAG-2 is dispensable for recombination(37,38). Although dispensable for recombination, the C-terminus of RAG-2 is essential for efficient V_H to DJ_H rearrangement at the immunoglobulin heavy chain(IgH) locus(52). Other domains not included in the core consist of a non-canonical plant homeodomain (PHD), an inhibitory domain(ID), a degradation domain, and a nuclear localization signal (NLS) (8,17,53,54).

The PHD of RAG-2 binds to the epigenetic mark histone 3 trimethylated at lysine 4 (H3K4me3)(7,17). This binding of the PHD to H3K4me3 is necessary for efficient antigen receptor rearrangement. Crystal structures of the RAG-2 PHD alone, and the RAG-2 PHD complexed with modified histone 3 peptides identified the PHD as capable of recognizing two distinct histone modifications, H3K4me2 and H3K4me3(8). The RAG-2 PHD is different from canonical PHDs because the RAG-2 PHD contains a Tyr instead of an arginine. Usually PHD binding to H3K4me3 is inhibited by the demethylation of arginine, but the RAG-2 PHD results in improved binding to H3K4me3 by the demethylation of arginine(8). Additionally, mutations in this region are linked to the pathogenesis of severe combined immunodeficiency (SCID) and Omenn syndrome(55,56).

The presence of an inhibitory domain was suggested by the paradoxical evidence that mutation of W453 in the PHD is more detrimental to recombination than the removal of the whole non-core domain(17). The inhibitory domain spans from residues 352-405 and is highly acidic (17,54). Disruption of the inhibitory domain by neutralization of the acidic residues to alanines, stimulates V(D)J recombination on both extrachromosomal substrates and at endogenous antigen receptor loci(54). Additionally, this disruption permits association of RAG with the IgH locus in the absence of H3K4me3 binding.

Other important regulatory domains in the non-core of RAG-2 are the degradation domain and the NLS. V(D)J recombination is temporally controlled through the periodic degradation of RAG-2. RAG-2 is phosphorylated at T490 by CyclinA/CDK2 at the G1-to-S cell cycle transition leading to subsequent ubiquitylation by the Skp2-SCF E3 ubiquitin ligase and proteasomal degradation(57-60). The periodic destruction of RAG-2 links V(D)J recombination to the cell cycle by restricting RAG activity to G0/G1(61). This coupling of V(D)J recombination to the cell cycle suppresses genomic instability and lymphoid tumorigenesis(62). In addition to the modes of regulation imparted by these non-core domains, the NLS is required for efficient nuclear localization of RAG-2 in cycling cells. RAG-2 is imported into the nucleus via binding of Importin 5 to residues 499-508 of RAG-2(63).

While the core domains of RAG-1 and RAG-2 are the minimal regions needed for DNA cleavage in vitro, these truncated proteins still exhibit reduced recombination levels in comparison to their full length counterparts(64). Of particular note, truncated RAG proteins generate 10 fold higher levels of signal ends, suggesting that the dispensable regions of RAG are involved in the proper processing of recombination intermediates(64). Mice containing core RAG-2 exhibit significant in vivo function, but have a reduced number of total B and T cells, indicating impaired lymphocyte development at the lymphocyte progenitor stages(65,66). These mice also exhibit defective ordered rearrangements, in which the numbers of V-to-D rearrangements within tri-partite antigen receptor loci are elevated and there is a defect in proper V_H-to-DJ_H recombination at the IgH locus (67,68). Mice expressing core RAG-1 and core RAG-2 exhibit an increase in aberrant V(D)J recombination and an increase in joins between signal and coding ends, termed hybrid V(D)J joins, suggesting that the non-core domains play a role in generating proper rearrangements (69,70).

Mechanism of V(D)J recombination

V(D)J recombination is the process by which functional antigen receptor genes are assembled from V, D and J gene segments. Each V, D and J gene segment family has multiple members, which can be combined with other gene segments in a different family to produce a 3.5 $x10^6$ potential V(D)J and VJ combinations(71). This combinatorial diversity, in combination with junctional diversity introduced by nucleotide loss and N and P nucleotide insertions, allows the immune system to recognize and respond to nearly any epitope with 10^{11} possible receptors(71). In brief, V(D)J recombination involves both the recognition of RSSs flanking antigen receptor gene segments by RAG-1 and recognition of H3K4me3 by RAG-2. After recognition of an RSS, RAG synapses two RSSs and induces a nick at the 5' end of the first heptamer. This is followed by a transesterification reaction which produces a hairpin at each coding end and a blunt, 5' phosphorylated break at each signal end. The coding joint hairpins are subsequently opened by Artemis and then joined by the classical non-homologous end joining machinery.

The first step in DNA cleavage requires RAG-1 binding of DNA. RAG-1 binds RSS DNA with 3-5 fold preference over a random sequence DNA, whereas RAG-2 doesn't exhibit direct binding of DNA(72). However, both RAG-1 and RAG-2 are needed for effective discrimination between an RSS and nonspecific sequences, as the RAG-1/RAG-2/RSS complex is more stable and specific than the RAG-1/RSS complex alone.

Cell free reactions followed by in vitro studies using purified proteins demonstrated that RAG-1 and RAG-2 are sufficient to perform DNA cleavage. In vitro coupled cleavage assays using radiolabeled DNA substrates and purified proteins revealed that upon synapsis of two RSSs,

nicking occurs prior to hairpin formation and that both RAG-1 and RAG-2 are required for nick and hairpin formation(73,74). RAG-1 introduces this nick at the junction of an RSS heptamer and the coding sequence; subsequent transesterification results in the formation of hairpins on the coding ends and blunt breaks on the signal ends(73,75). The coupled cleavage reaction requires two RSSs with spacers of 12 and 23 bp, a restriction known as the 12/23 rule(76).

This 12/23 rule is established prior to DNA cleavage(77). The spacer ensures that there is an appropriate distance between the heptamer and nonamer of the RSS, and also regulates activity by providing additional RAG:RSS interaction surfaces(78). The 12/23 rule not only imparts specificity but also is involved in recruitment of RAG-1. RAG-1 is recruited most efficiently to substrates with a 12/23 RSS pair(79). Heptamer, spacer, and nonamer elements are all required for coupled cleavage in vivo(80). Within the RSS, the nonamer is important for initial binding and some coordination of nick and hairpin formation(81). Mutation of this nonamer leads to reduced stability of RAG with substrate DNA(82). In contrast, the heptamer is important for hairpin formation and mutation of the heptamer does not affect the stability of the RAG-RSS complex(82). The heptamer, specifically the three nucleotides (CAC) adjacent to the coding sequence, is essential for V(D)J recombination(83).

Additional junctional diversity is introduced into the coding joint by the location of hairpin opening, potential insertions/deletions by the NHEJ machinery, and the addition of nontemplated nucleotides (N regions) by terminal deoxynucleotidyl transferase (TdT). The hairpinned coding ends are opened by Artemis (84,85). Coding joints are then formed by the classical NHEJ machinery. XRCC4 and Ku are required for both DSB repair and V(D)J recombination. Ku is a

heterodimer of 70 and 86kDA subunits, which binds to DNA broken ends (85). Ku is required for efficient V(D)J recombination as Ku86 deficient mice are defective in forming coding or signal joints and exhibit coding joints without N regions (86). After Ku binding, DNA-PKcs and TdT are recruited to the DNA ends and DNA-PKcs can phosphorylate its substrates(87-89). Additional diversity is introduced at the coding junction by terminal deoxynucleotidyltransferase (TdT) which inserts nontemplated (N) nucleotides(90-92) and palindromic (P) nucleotides created through the asymmetric opening of the hairpin(93). The final step of classical NHEJ involves the ligation of DNA by DNA ligase IV(87) to form a coding joint and a signal joint.

3-Dimensional structure of RAG-1 and RAG-2

The first crystal structure of RAG-1 and RAG-2 revealed that the V(D)J recombinase adopts a 'Y-shaped' conformation consisting of a heterotetrameric complex with two subunits of RAG-1 and two subunits of RAG-2, however DNA was not visible in this structure(5). Initial evidence for this model came from coimmunoprecipitation experiments demonstrating that RAG-1 and RAG-2 associate with each other within a complex(94). RAG-1 residues 504-1008 were sufficient to co-immunoprecipitate with RAG-2(95). Initial insights into the stoichiometry of the complex were incorrect, but further experiments demonstrated that core RAG-1 exists as a dimer when free in solution and this dimer was the minimal species bound to RSS(96). In addition to RAG-1 and RAG-2, the high mobility group protein HMG2 can be stably incorporated into the RAG-1/RSS complex. In the absence of RAG-2, HMG2 can increase the affinity of RAG-1 for RSS (96,97). The notion that HMG proteins can stabilize the RAG complex is supported by the evidence that HMG1 has greater affinity for RAG-1-DNA complexes than for RAG-1 or DNA alone(98).

Prior to the 2015 crystal structure, a number of structural insights came from smaller, domain-specific studies. One such study demonstrated that the nonamer binding domain of RAG-1 forms a dimer that can bind and synapse two nonamer sequences(99). Fluorescence resonance emission transfer (FRET) experiments demonstrated that the 23-RSS is bent to facilitate the recognition of the heptamer and nonamer by RAG(100). RAG-1 alone can bind the nonamer, but RAG-2 is required for heptamer binding because it stabilizes RAG-1-heptamer interactions(101,102). Crystal and cryo-EM structures of RAG-1 and RAG-2 with DNA in the pre-reaction and hairpin forming complexes revealed that coding-flank DNAs extensively move and deform for nicking and hairpin formation(103). Recent cryo-EM structural studies in zebrafish RAG have detected DNA melting at the heptamer (103), in agreement with the results of earlier DNA modification interference experiments (101). Heptamer melting is seen to place the scissile phosphate in the RAG-1 active site(104).

Subsequent crystallographic analysis of core RAG-1 and core-RAG-2 in a signal-end complex confirmed many of the conclusions drawn from early structural insights. In particular the new data indicated that in a signal-end complex, RAG forms a 'Y' shaped structure in which the RAG-1 dimer forming the majority of the 'Y' with RAG-2 is located at the tip of each arm of the 'Y' (5). RAG-1 contains an RNase H fold, in common with other DDE transposases. This structure revealed that RAG-2 contacts RAG-1 near the active site (5) but did not provide information regarding the non-core regions, which were absent.

Several structures of the RAG-2 PHD, alone or in complex with histone 3 peptides, have been determined(7,8). Unlike other PHDs that bind histone H3K4me3, the RAG-2 PHD in complex with an H3K4me3 peptide lacks the canonical aromatic cage at the H3K4me3 binding site, which instead consists of an aromatic channel comprised of the highly conserved residues Y415, M443, and W453(7). Additionally, structural analysis showed that five amino acid residues that are mutated in primary immunodeficiencies participate in histone H3 recognition(8). These structural studies suggested a connection between histone methylation and V(D)J recombination.

The most recent structural information comes from a cryo-EM structure of mouse RAG-1 (residues 265-1040) and RAG-2 (residues 1-520) in complex with various DNA substrates(105). Unlike most DDE transposases, RAG first cleaves the upper strand at the 5' boundary of the RSS with the coding sequence. In the pre-recombination complex (PRC) the RAG active site remains partially constructed and inactive, with only two of the acidic residues in close proximity(105). The active site then adopts two different configurations to performing nicking and hairpinning reactions in the nick forming complex (NFC) and hairpin forming complex (HFC) respectively. The PRC and NFC are similar from RAG-2 at the tip of the 'Y' structure through the NBD of RAG-1 in the 'Y' stem. However the NFC has ZnH2 domains that open outward upon binding to DNA in the PRC and then move inward when creating the NFC(105). This movement is also reflected in changes in contacts to the heptamer from PRC to NFC. In the NFC, a DNA zipper is formed which enables DNA unwinding and places the scissile phosphate in the active site(105). After hydrolysis, the 3'OH remains in the active site and serves as the nucleophile for producing a hairpin.

After DNA cleavage, a stable post-cleavage complex is formed between synapsed recombination signal sequences, RAG-1, RAG-2, HMG-1/HMG-2, and components of DNA-dependent protein kinase(106). A mutant form of RAG-1(S723C) is proficient for DNA cleavage but has defects in post-cleavage complex formation and end joining in vitro. A mouse model expressing the RAG-1(S723C) mutation is a hypomorph which exhibits aberrant DNA double-stranded breaks, impaired lymphocyte development and decreased V(D)J joining(107). Early structural analyses of the post-cleavage complex indicated that the signal-end complex contains two protomers of RAG-1 and two protomers of RAG-2, a structure that was later confirmed(5,18). This complex has two fold symmetry and the N termini of RAG-1 and RAG-2 are positioned at opposing ends of the complex(18).

Regulation of V(D)J recombination

The DNA breaks introduced by RAG are potentially deleterious, and V(D)J recombination is tightly regulated to mitigate genomic instability. Regulation of this process occurs in lineage-specific and developmental stage-specific fashion. One layer of this regulation involves the controlled expression of RAG-1 and RAG-2.

V(D)J recombination is restricted to the B and T lymphoid lineages and this is enforced in part by restricted patterns of RAG-1 and RAG-2 expression (108). RAG-1 and RAG-2 are exclusively expressed in the lymphoid lineage and in a developmentally regulated fashion. For example within the T lineage, RAG-1 and RAG-2 are both expressed in immature $CD4^+CD8^+$ (double positive) T cells, but not in more mature $CD4^+CD8^-$ or $CD4^-CD8^+$(single positive) subpopulations (109). The promoter of RAG-2 is lymphoid specific, but is regulated differently in B cells and T cells (110). In B cells, the B-cell-specific transcription factor BSAP/Pax-5 can bind to the promoter, while in T cells c-Myb regulates the RAG-2 promoter (111). In B cells, a conserved transcriptional enhancer termed Erag is essential for expression off of RAG-1 and RAG-2 promoters(68). C-Myb negatively regulates Foxo1 expression, an activator of RAG(112), and also occupies the Erag enhancer(112)..

Within a lineage, RAG is also developmentally controlled. At the immunoglobulin locus, heavy chain rearrangement precedes light chain rearrangement, and within the heavy chain locus D_H-to-J_H rearrangement occurs before V_H-to-DJ_H rearrangement (113). RAG transcript levels are high in pro-B and pre-B-I cells, in which in D_H-to-J_H and V_H-to-DJ_H rearrangement occurs, lower in

large pre-B-II cells and high again in small preB-II cells, which undergo V_L-to-J_L rearrangement(114).

Another mode of developmental regulation is imposed by chromatin remodeling. Alterations to nucleosome structure by chromatin remodelers, such as SWI/SNF that increase histone acetylation, can increase the accessibility of nucleosomal DNA to cleavage by RAG (115). Regulated accessibility of individual antigen receptor loci by histone marks helps to control the process of V(D)J recombination. Epigenetic alterations of chromatin at antigen receptor loci are regulated in a locus, lineage, and stage specific manner during B cell development(116).

Beyond lineage and developmental regulation, V(D)J recombination is restricted to antigen receptor loci at the level of RAG-1 and RAG-2 discrimination. In vivo, RAG-1 binds only at regions containing RSSs while RAG-2 binds to thousands of genomic sites containing histone 3 trimethylated at lysine 4 (H3K4me3)(117). I will expand upon epigenetic regulation of V(D)J recombination in the next section. Moreover, recombination appears to be topologically restricted to chromatin loops, which limits potential recombination events to those containing gene segments within special proximity to each other(118). This mode of restriction is discussed more fully in Chapter 2 of this dissertation.

Transcription and epigenetic control of the RAG recombinase

Transcription and V(D)J recombination are tightly linked(10). Transcription of unrearranged V_H gene segments is positively correlated with the targeting of these segments for V(D)J recombination(10). This has suggested a model, termed the accessibility model, in which transcription establishes an open chromatin environment that provides accessibility to the V(D)J

recombinase. This model suggests that the lineage- and developmental stage specificity of V(D)J recombination are established by distinct patterns of transcription. A causal relationship between transcription and recombination was suggested by the demonstration that transcription of neighboring genes enhances substrate recombination efficiency in a pre-B cell model with stably integrated recombination substrates(119). In agreement with the accessibility model, rearranged antigen receptor loci exhibit greater sensitivity to DNase1 digestion compared to unrearranged loci(120).

Transcriptional control elements such as enhancers and promoters were identified within antigen receptor loci and upon recruitment of transcription factors, these elements could activate transcription and recombination(121). Further insights into these accessibility control elements came from studies examining the intronic IgH enhancer (Emu). Mutational analysis of transcriptional control elements showed that Emu is necessary for V(D)J recombination in developing B cells and IgH locus transcription in mature B cells(122-124). Enhancer deletion experiments at other antigen receptor loci yielded similar results, pointing to the importance of enhancers in regulating recombination(122,123,125-131). On a more local scale, transcriptional promoters can influence recombination. Deletion studies in the TRA and TRB loci demonstrated that removal of promoters can decrease recombination at proximal gene segments(126,132-134). Testing of the accessibility hypothesis led to the conclusions that enhancers control RAG-1 binding globally, that promoters and transcription can direct RAG-1 binding locally, and that RAG-1 can be targeted to DNA in the absence of RAG-2 (135).

The lineage-specificity and temporal ordering of antigen receptor loci rearrangement is reflected by the relative chromatin accessibility of the RSS(136,137). Chromatin states can inhibit recombination in vivo (138-141). Studies using chromatinized or naked DNA substrates demonstrated that ATP-dependent remodeling of chromatin can result in increased DNA cleavage by RAG(142). Accessibility of antigen receptor loci is also dependent on nucleosome

distribution. At the immunoglobulin heavy chain (IgH) locus in pro-B cells, a nucleosome is positioned over each IgH variable coding gene segment which places the RSS in a position accessible to RAG(143). In contrast, in mouse embryonic fibroblasts, the IgH coding gene segments are depleted of nucleosomes, which then result in nucleosomes covering the RSS making the DNA inaccessible to RAG. In a cell-free system, recruitment of SWI/SNF chromatin-remodelers to an RSS increases coupled cleavage by RAG demonstrating a role for chromatin remodeling in V(D)J recombination(144). Additionally, depletion of SWI/SNF in pro-B cells prevents antisense transcription at the IgH locus(145). Thus, changes in nucleosome positioning, in addition to chromatin modifications, may play a role in controlling V(D)J recombination during development.

Within the IGH locus epigenetic marks help to constrain V_H recombination to already recombined DJ_H junctions(137). These DJ_H junctions are marked with active histone marks, bind to RAG, and become nuclease sensitive(15). Analyses in RAG-deficient pro-B cells demonstrated that only the 5' and 3' most D_H gene segments, and the J_H segments have hallmarks of active chromatin. Additionally, the J_H segments have the highest density of RAG proteins in the germline. H3 and H4 acetylation and H3K4dimethylation (H3K4me2) are restricted to the 5' and 3' most D_H segments called DFL16.1 and DQ52, while D_H gene segments in between are marked by the heterochromatic modification H3K9me2. This epigenetic pattern is reflected in recombination, where the DFL16.1 and DQ52 gene segment recombine more frequently than other D segments(11,146). Upon rearrangement, RAG becomes redistributed toward the DJ_H junctions(15). Chromatin immunoprecipitation analyses in Emu deficient cell lines demonstrated that Emu is necessary for the deposition of these active histone marks at the IgH locus(15). Both the germline and DJ_H rearranged loci lacked the active histone marks H3K4me3 and H3K9ac(15). These data further support the accessibility model and provide insight into the ordered rearrangement at the IgH locus.

The finding that RAG-2 binds to H3K4me3, and that this binding is essential for efficient rearrangement, provided a direct mechanistic link between transcription and V(D)J recombination. RAG-2 contains a PHD finger that specifically recognizes H3K4me3. Mutations that inhibit RAG-2 recognition of H3K4me3 decreased V(D)J recombination in vivo(6,7). A conserved tryptophan residue W453 is a key structural component of the H3K4me3 binding surface and is required for RAG-2 recognition of H3K4me3(6,7). Mutations of W453, specifically W453R, is associated with the primary immunodeficiencies termed Omenn syndrome (55). The pattern of RAG-2 binding to chromatin across the IgH locus is positively correlated with the distribution of H3K4me3(6).

The requirement for H3K4me3 recognition can be bypassed by mutations within an inhibitory domain in RAG-2, whose disruption is associated with constitutive increases in RSS binding affinity, catalytic rate and recombination frequency, mimicking the stimulatory effects of H3K4me3 (17). Association of RAG-1 and RAG-2 with the IgH locus in B cell progenitors is dependent on recognition of H3K4me3 by the PHD (117,147) and disruption of the inhibitory domain permits association of RAG with the IgH locus in the absence of H3K4me3 binding (54). Thus, the inhibitory domain blocks access of RAG to the IgH locus except when H3K4me3 is engaged by RAG-2(17,54). RAG-1 binding is driven by both H3K4me3 via RAG-2 and also by H3K27Ac through the non-core portions of RAG-1(148). Not only does H3K4me3 bind to RAG-2, but it allosterically activates RAG and induces conformational changes in RAG-1(6,17,20). Our understanding of the coupling between transcription and antigen receptor gene rearrangement has been advanced by the finding that epigenetic marks deposited during transcription can allosterically activate RAG and stimulate recombination.

Approach

Taken together these observations suggest the following model for association of RAG with antigen receptor loci: (1) in the absence of H3K4me3, RAG has low affinity for the RSS; (2) when the RAG-2 PHD finger is engaged by H3K4me3, an allosteric signal is propagated to RAG-1, which acquires high affinity for the RSS; and (3) the binding of RAG to a nearby RSS stabilizes its association with the locus. As a test of this model, we sought to identify mutations in the RAG-2 PHD finger that separate binding of H3K4me3 from allosteric activation of RAG-1. Here we employ a strategy of phylogenetic substitution to identify amino acid residues within the PHD finger, but distinct from the H3K4me3 binding site, that function in transmission of an allosteric signal to RAG-1.

Results

It has remained unclear how information regarding the binding of H3K4me3 by RAG-2 is transmitted to RAG-1. We hypothesized that such information is conveyed through a path involving amino acid residues within the PHD finger other than those that engage H3K4me3. To test this hypothesis, we sought to identify separation-of-function mutations in the RAG-2 PHD finger that impair V(D)J recombination but spare binding to H3K4me3.

Phylogenetic substitution separates H3K4me3 binding from allosteric activation in vitro

To decrease the complexity of the search we employed a phylogenetic approach, reasoning that RAG-2 PHD fingers from species distantly related to the mouse might retain the ability to bind H3K4me3 but be sufficiently different in sequence so that transmission of the allosteric signal to mouse RAG-1 might be impaired. RAG-2 of the brownbanded bamboo shark, *C. punctatum*, is phylogenetically distant from that of the mouse (149). In the interval spanning residues 414 through 485 the RAG-2 PHD finger of *M. musculus* differs from that of *C. punctatum* at 25 positions (Fig 1A). When mapped onto a crystal structure of the mouse PHD finger (7), these differences largely reside opposite to the H3K4me3 binding pocket (Fig. 1B) and are potentially free to participate in an allosteric interface.

The similarity of the *M. musculus* and *C. punctatum* PHD fingers in the region of the H3K4me3 binding site suggested that the shark PHD finger would also be capable of binding H3K4me3. To test this, we constructed a RAG-2 chimera in which the mouse PHD finger was replaced by the PHD finger from *C. punctatum*. Lysates of HEK293T cells expressing wild-type, murine RAG-2 or the murine RAG-2 bearing the *C. punctatum* PHD finger were incubated with a bead-bound peptide corresponding to residues 1 - 21 of histone H3 and trimethylated at lysine 4 (H3K4me3). Both wild-type murine RAG-2 and the mouse-shark RAG-2 chimera bound to the

H3K4me3 peptide (Fig. 1C, right, lanes 1 and 3). As expected, a W453A mutation in the H3K4me3 binding pocket of mouse RAG-2 abolished binding to H3K4me3 (Fig. 1C, right, lane 2). In addition, the KMT2D lysine methyltransferase, which carries multiple PHD fingers with specificity distinct from that of RAG-2, also failed to bind H3K4me3 (Fig. 1C, right, lane 4).

We next asked whether the *C. punctatum* PHD finger could replace its murine counterpart in an assay for allosteric activation of coupled DNA cleavage by RAG. The version of RAG-1 used in these assays, cR1ct, lacks the amino-terminal non-core region and is more soluble than the wild-type protein but remains responsive to H3K4me3 (17,150). Full-length, wild-type mouse RAG-2 and chimeric RAG-2 bearing the *C. punctatum* PHD finger, both tagged at the amino terminus with MBP, were coexpressed individually with MBP-tagged cR1ct and RAG tetramers were purified by a protocol that removes endogenous H3K4me3 (20). Equivalent amounts of each RAG tetrameric complex were assayed in vitro for coupled cleavage of a radiolabeled 12-RSS substrate in the presence of an unlabeled 23-RSS substrate and increasing amounts of a histone H3-derived peptide containing trimethylated lysine 4 (H3K4me3) or unmethylated lysine 4 (H3K4me0). As previously reported (17,19,150), formation of nicked product was stimulated in a dose-dependent manner by H3K4me3, but not by H3K4me0 (Fig 1D). In contrast, the chimeric version of RAG-2 failed to be stimulated by H3K4me3 (Fig. 1D). Similar results were obtained with independent protein preparations. Taken together, these observations indicated that replacement of the mouse RAG-2 PHD finger with the PHD finger from *C. punctatum* produces a protein in which the H3K4me3-binding is retained while the ability to support allosteric activation of RAG-1 is lost.

A RAG-2 PHD mutant that binds H3K4me3 without supporting V(D)J recombination in vivo

We went on to test whether the binding of H3K4me3 could be separated from the ability of RAG-2 to support V(D)J recombination in vivo. An initial assay for recombination of an extrachromosomal substrate (17,54,151) suggested that the mouse-shark chimeric RAG-2

(PHD_{PUNC}) was impaired in its ability to support V(D)J recombination, as its activity was similar to that of RAG-2(W453A), which cannot bind H3K4me3, and of a chimeric RAG-2 mutant bearing the corresponding mutation, W455A in the *C. punctatum* PHD finger (Supplemental Fig. 1).

We proceeded to a more extensive analysis, with the goal of identifying mutations that impair V(D)J recombination while disrupting neither protein expression nor the ability to bind H3K4me3. Cassettes encoding wild-type RAG-2 and a series of variants, fused at the amino terminus to the FLAG epitope and at the carboxyl terminus to poly-histidine and myc tags (Fig. 2A), were introduced into the retroviral vector pCLIP2a (152). These included RAG-2(PHD_{PUNC}), the chimera bearing the *C. punctatum* PHD finger in the context of mouse RAG-2. As negative controls we tested RAG-2(W453A) and RAG-2(PHD_{PUNC}, W455A) (Fig. 2A), in which an invariant tryptophan residue involved in H3K4me3 binding was mutated to alanine. We have previously shown that disruption of the RAG-2 inhibitory domain permits V(D)J recombination to override the requirement for recognition of H3K4me3 (17,54). It was therefore also of interest to assay RAG-2(D/E352-405A) and RAG-2(D/E352-405A, PHD_{PUNC}) (Fig. 2A), in which the acidic residues within the inhibitory domain were mutated to alanine. We constructed an additional set of mutants in which groups of amino acid residues in the *C. punctatum* PHD finger were mutated to the corresponding residues in the mouse. Using the three-dimensional structure of the mouse PHD finger as a guide, we assigned these sequence differences to five groups, each of whose residues are closely clustered on the surface opposite the H3K4me3 binding site (Fig. 2B and 2C), reasoning that one or more alterations on this surface might disrupt recombination but spare binding to H3K4me3. These spatially clustered mutations, as depicted in Fig. 2B, were termed Group 1 (K418T, S420C, S421P, A422T, N424D), Group 2 (M425V, N426D, V427I, I429T, E431V, Y433F), Group3 (N460D, A462E, S464R, Q465T, Q468H, F469L), Group 4 (Q471E, E472G, N473S, T474N) and Group 5 (F445Y, S448H, F477Y, Y482V), where the numbering follows that of the mouse sequence.

To assay recombination quantitatively we employed a fluorescent assay for recombination of an integrated substrate (153). Retroviral constructs encoding the RAG-2 variants described in Fig. 2A, as well as constructs encoding the clustered shark-to-mouse mutants, were used to infect the cell line R2K3, a RAG-2-deficient, Abelson murine leukemia virus (AbMuLV)-transformed B progenitor (153). R2K3 contains an integrated substrate for V(D)J recombination whose inversional rearrangement results in the expression of GFP. Following transduction of RAG-2 recombination is induced by STI-571; rearrangement of the integrated substrate is detected by flow cytometry. Cells transduced with wild-type RAG-2 exhibited increasing levels of GFP fluorescence, approaching a recombination frequency of 80 percent, over a period of at least 96 hr following arrest with STI-571 (a representative analysis at 96 hr is shown in Fig. 3A; the results of three technical replicate assays at 48 hr and 96 hr are shown in Fig. 3B). As expected, recombination was impaired by the W453A mutation, which disrupts binding of H3K4me3 (Fig. 3A and 3B). We also observed little recombination in cells transduced with the RAG-2(PHD_{PUNC}) chimera or with RAG-2($PHD_{PUNC,}$W455A) (Fig. 3A and 3B). Notably, the RAG-2(D/E352-405A, PHD_{PUNC}) chimera, in which the inhibitory domain is neutralized, supported recombination to a level near to that of wild-type RAG-2 or RAG-2(D/E352-405A) (Fig. 3A and 3B, $ID_{NEUT}PHD_{PUNC}$), indicating that the chimeric protein can interact functionally with RAG-1 to support recombination when the requirement for H3K4me3 binding is bypassed. Among the back-mutated mutants, RAG-2(PHD_{PUNC}, Group1) and RAG-2(PHD_{PUNC}, Group4) showed sharply diminished recombination activity with respect to wild-type RAG-2, while RAG-2(PHD_{PUNC}, Group 2), RAG-2(PHD_{PUNC}, Group 3) and RAG-2(PHD_{PUNC}, Group5) supported recombination at frequencies of about 38 to 75 percent that of wild-type (Fig. 3A and 3B).

Because any impairment of recombination could result from decreased protein accumulation, we examined expression of each of the RAG-2 variants at 96 hr after arrest by STI-571, with the goal of identifying mutant proteins that are robustly expressed but fail to support efficient V(D)J recombination. Accumulation of RAG-2(PHD_{PUNC}) was reduced to less than one tenth that of

wild-type RAG-2 (Fig. 3C and Supplemental Fig. 2) and could not be rescued by overexpression of RAG-1 in a cotransfection assay (Supplemental Fig. 3). Of the additional mutant proteins that exhibited impaired recombination in the R2K3 assay, only RAG-2(PHD_{PUNC}, Group 1) was expressed at a level similar to those RAG-2 variants that retained recombination activity (Fig. 3C and Supplemental Fig. 2). Thus, the Group 1 mutation permitted near normal accumulation of the RAG-2(PHD_{PUNC}) chimeric protein but was associated with a large reduction in recombination activity.

To confirm the effects of the RAG-2 mutations described above on V(D)J recombination, the R2K3 rearrangement assay was performed in biological triplicate, each biological assay consisting of three technical replicates (Fig. 3D). Of note, cells transduced with RAG-2(PHD_{PUNC}, Group 1) exhibited a large and statistically significant decrease in recombination frequency in comparison to cells transduced with wild-type RAG-2 or RAG-2(PHD_{PUNC}, Group 2) (Fig. 3D, Group 1), while neutralization of the inhibitory domain robustly conferred recombination activity on the RAG-2(PHD_{PUNC}) chimera (Fig. 3D, ID_{NEUT}, PHD_{PUNC}).

A back mutation from shark to mouse sequence in the RAG-2 PHD that restores V(D)J recombination activity

We considered that the impairment of recombination activity observed for the RAG-2(PHD_{PUNC}, Group 1) mutant might result from disruption of intramolecular interactions that propagate the allosteric signal initiated by the binding of RAG-2 to H3K4me3. To test this interpretation we asked: (1) whether the defect in recombination observed for RAG-2(PHD_{PUNC}, Group 1) can be reversed by an additional back-mutation within the *C. punctatum* PHD finger; and (2) whether RAG-2(PHD_{PUNC}, Group 1) retains the ability to bind H3K4me3.

Because RAG-2(PHD_{PUNC}, Group 2) robustly supported V(D)J recombination, the Group 2 back-mutation was an attractive candidate for rescue of RAG-2(PHD_{PUNC}, Group 1). We observed that

a mouse-shark RAG-2 chimera bearing combined Group 1 and Group 2 mutations exhibited a robust increase in recombination frequency, relative to the Group 1 mutant (Fig. 4A and 4B, compare Group 1 to Groups 1+2). In this assay the expression levels of RAG-2(PHD_{PUNC}, Groups1+2), RAG-2(PHD_{PUNC}, Group1) and wild-type RAG-2 differed by about two-fold or less (Fig. 4C). Taken together, the results described above indicate that the Group 1 mutation enhances accumulation of the RAG-2(PHD_{PUNC}) chimera without permitting recombination, and that additional back-mutation of the Group 2 cluster confers recombination activity on the Group 1 mutant. Consistent results were obtained with a similar recombination assay in which RAG-2 variants were transduced using the pCST virus (Supplemental Fig. 4).

To establish that RAG-2(PHD_{PUNC}, Group 1) is indeed a separation-of-function mutant, we tested its ability to bind H3K4me3. Wild-type mouse RAG-2, RAG-2(W453A), RAG-2(PHD_{PUNC}), RAG-2(PHD_{PUNC}, Group 1), RAG-2(PHD_{PUNC}, Group 2) and RAG-2(PHD_{PUNC}, Groups 1+2) were expressed as MBP fusions in HEK293 cells. Lysates were incubated with beads coated with the H3K4me3 peptide or the corresponding unmethylated peptide (H3K4me0). Bead-bound protein was detected by immunoblotting for MBP. As expected, wild-type RAG-2 was capable of binding to H3K4me3 but not to H3K4me0, while RAG-2(W453A) bound neither peptide (Fig. 4D). Consistent with results presented above, RAG-2(PHD_{PUNC}) was able to bind H3K4me3. Importantly, RAG-2(PHD_{PUNC}, Group 1), RAG-2(PHD_{PUNC}, Group 2) and RAG-2(PHD_{PUNC}, Group 1) each exhibited binding to the H3K4me3 peptide, but not to H3K4me0 (Fig 4D). Taken together the properties of the RAG-2(PHD_{PUNC}, Group 1) mutant indicate that the ability of the RAG-2 PHD finger to bind H3K4me3 can be separated from its ability to support V(D)J recombination.

Discussion

Binding of H3K4me3 to the RAG-2 PHD finger alters the conformation of RAG-1 and this is accompanied by enhancements of affinity for substrate and catalytic rate (6,7,17,19,20) These observations have suggested an allosteric model for activation of RAG. According to such a model, binding of H3K4me3 by the RAG-2 PHD finger would be necessary but insufficient for activation of RAG-1. The observations described here provide additional support for an allosteric model by demonstrating that activation of RAG-1 requires one or more functions of the RAG-2 PHD finger in addition to engagement of H3K4me3.

An initial indication that H3K4me3 binding is insufficient for activation of RAG was provided by the RAG-2(PHD_{PUNC}), which retains the ability to bind H3K4me3 but was unresponsive to allosteric activation in vitro by an H3K4me3 peptide under conditions in which equivalent amounts of wild-type RAG and chimeric mutant were compared. Although this observation is consistent with allostery, we considered a more trivial explanation for loss of activity, namely, that RAG-2(PHD_{PUNC}) may have undergone a structural alteration that renders it unable to support DNA cleavage, for example, by disrupting association with RAG-1. This is rendered unlikely by the fact that disruption of the inhibitory domain restores the ability of the RAG-2(PHD_{PUNC}) chimera to support recombination. We therefore favor the interpretation that the PHD finger of *C. punctatum* is unable to transmit an activating allosteric signal to mouse RAG-1 upon binding of H3K4me3.

Although RAG-2(PHD_{PUNC}) satisfied biochemical criteria for a separation-of-function mutant, its poor expression in vivo precluded our ability to assess its function in V(D)J recombination. By systematic reversion of RAG-2(PHD_{PUNC}) residues specific to *C. punctatum* to those of the mouse, we identified three clustered mutations, RAG-2(PHD_{PUNC}, Group2), RAG-2(PHD_{PUNC}, Group3) and

RAG-2(PHD_{PUNC}, Group5), which were able to confer recombination activity. Because each of these mutations also restored protein accumulation to an amount similar to that of wild-type RAG-2, their stimulatory effect on recombination may have been a consequence of increased expression. Alternatively, or in addition, these mutations may have effected an increase in the affinity of the PHD finger for H3K4me3 or enhanced the efficiency with which information regarding the engagement of H3K4me3 is transmitted to RAG-1. In this regard it should be noted that the clustered Group 3 and Group 5 mutations reside near the H3K4me3 binding site.

The RAG-2(PHD_{PUNC}, Group1) mutant, like RAG-2(PHD_{PUNC}, Group2), RAG-2(PHD_{PUNC}, Group3) and RAG-2(PHD_{PUNC}, Group5), was expressed robustly in vivo. Unlike RAG-2(PHD_{PUNC}, Group2), RAG-2(PHD_{PUNC}, Group3) and RAG-2(PHD_{PUNC}, Group5), RAG-2(PHD_{PUNC}, Group1) exhibited a large and significant reduction in the ability to support V(D)J recombination, even though it retained the ability to bind H3K4me3. Thus, with respect to recombination RAG-2(PHD_{PUNC}, Group1) represents a separation-of-function mutant. We considered the possibility that RAG-2(PHD_{PUNC}, Group1) may have suffered an overall, debilitating structural lesion but consider this unlikely because RAG-2(PHD_{PUNC}, Groups1+2), in which the Group 1 and Group 2 mutations are combined, exhibited full function. Lastly, we considered that RAG-2(PHD_{PUNC}, Group1) may have exhibited a loss of function because of an inability to interact with an essential factor other than RAG-1. This is unlikely because allosteric activation of coupled RSS cleavage does not require components other than RAG-1, RAG-2 and H3K4me3 peptide.

Available RAG structures do not include the PHD finger and therefore do not provide information regarding its interactions (5,154,155). A recent study (156) provides evidence for an interaction between RAG-1 and the dispensable region of RAG-2 but the structural basis for this interaction remains unexplored. The Group 2 mutation, which converted six residues in the *C. punctatum* PHD finger - M425, N426, V427, I429, E431 and Y433 - to the corresponding residues of the mouse was able to restore function to the RAG-2(PHD_{PUNC}, Group1) mutant in an

assay for V(D)J recombination. The ability of the Group 2 mutation to rescue RAG-2(PHD_{PUNC}, Group1) is not likely to be the result of effects on protein accumulation or H3K4me3 binding, although the latter possibility is not excluded by the available data. This result strongly suggests that in the mouse RAG-2 PHD finger one or more of the residues V425, D426, I427, T429, V431 and F433 participate in the relief of autoinhibition exerted by the RAG-2 inhibitory domain, either through transmission of the H3K4me3 allosteric signal or in some other, as yet undefined way. Because the mutational strategy employed here was restricted to amino acid sequence differences between *M. musculus* and *C. punctatum*, it remains likely that other residues contribute to the activation of RAG-1 in response to H3K4me3.

Although the PHD finger of RAG-2 from *C. punctatum* binds H3K4me3 it has not, to our knowledge, been determined whether the *C. punctatum* RAG ortholog is subject to allosteric activation by H3K4me3. Alignment of RAG-2 from M. musculus and C. punctatum does, however, reveal 43 percent sequence identity within the interval from residues 352 through 405 of the mouse, corresponding to the inhibitory domain (54) Moreover, the corresponding *C. punctatum* sequence, like that of the mouse, is highly acidic, suggesting the presence of an autoinhibitory function within RAG-2 of *C. punctatum* that may be relieved by H3K4me3.

The PHD finger of human RAG-2 is a frequent target of mutations associated with primary immunodeficiency diseases (157). Disease-associated missense mutations have been identified at thirteen sites within the PHD; mutations at eight of these residues have been found in patients with Omenn syndrome, a primary immunodeficiency disorder characterized by pathogenic oligoclonal T cell expansion and profound B lymphopenia (55,157-159) The remaining mutations are associated with severe combined immunodeficiency (SCID), atypical severe combined immunodeficiency or combined immunodeficiency with granulomas and/or

autoimmunity (157,159). Several of these mutations, including those at invariant residues C446 and C478, which participate in the formation of a zinc finger, are expected to disrupt the three-dimensional structure of the domain (7,55). Other mutations, including those at residues M443 and W453, abrogate binding of H3K4me3 (6,7). Thus, several RAG-2 mutations associated with primary immunodeficiency can be inferred to disrupt allosteric activation of RAG-1 by H3K4me3. Of the known disease-associated mutations in the RAG-2 PHD only one, N474S, is contained within one of the clustered mutations tested in this communication; all of the others affect residues that are conserved between *M. musculus* and *C. punctatum*. Notably, the N474S mutation does not impair the ability of RAG-2 to support V(D)J recombination (55,159) suggesting that in this case one or more other mutations are responsible for the disease phenotype. Pathogenic mutations within the RAG-2 PHD at sites other than N474 are associated with impaired V(D)J recombination activity (55,159), and it remains possible that one or more of these mutations interferes with allosteric transmission of the H3K4me3 signal to RAG-1.

Summary and Future Directions

V(D)J recombination is associated with gene transcription and epigenetic marks of active chromatin such as H3K4me3. Binding of H3K4me3 by the RAG-2 PHD results in conformational changes in RAG-1 and allosteric activation of RAG. In this chapter, we have demonstrated that H3K4me3 binding and transmission of the allosteric signal to RAG-1 are separable functions of the RAG-2 PHD finger. Additionally, we have begun to identify the path of allostery from the RAG-2 PHD finger to RAG-1.

To build upon this work, it would be interesting to further utilize the RAG-2(PHD_{PUNC}, group1) separation of function mutant to better define the allosteric surface responsible for transmitting allostery between RAG-1 and RAG-2. In particular, identifying individual residues within RAG-2(PHD_{PUNC}, group2) that are necessary for rescue of allostery would be insightful. One of the limitations of this dissertation is that it only examined residues that differed between *M. musculus* and *C. punctatum*. It remains possible that residues that are invariant between these two species may also play a role in forming this allosteric surface. Future work would need to be conducted to identify these residues. However, with the separation of function mutant in hand, we could easily further define which residues within group2 are responsible for the rescue of allosteric activation.

This separation of function mutant can be utilized to further define the mechanism behind this allosteric activation. It would be interesting to determine which residues of the putative allosteric surface, if any, make direct contact with RAG-1. Co-immunoprecipitation experiments utilizing our new RAG-2 mutants with RAG-1 could provide much needed mechanistic insight into this interaction. It would also be interesting to determine the binding constants for these mutants to H3K4me3 utilizing an EMSA. It is possible that restoration of allostery could be affected by the strength of binding to H3K4me3.

We also did not define which residues within RAG-1 may be involved in allosteric transmission. It would be interesting to utilize a similar phylogenetic substitution approach with RAG-1 to identify residues in RAG-1 that associate with the proposed allosteric surface of RAG-2. Initial insights into this could come from computational analyses such as covariation analyses. By

examining the co-evolution of residues in RAG-1 and RAG-2, we may be able to isolate patterns of co-evolution between the two proteins suggesting a potential interaction of those residues.

It remains possible that allosteric transmission through a PHD is a more generalizable phenomenon. Our insights could provide useful information for how PHDs in general are implicated in allostery. Additional experiments utilizing PHD fingers from various species and different proteins would be needed to support this hypothesis.

Overall this dissertation provides new mechanistic insight into how V(D)J recombination is regulated at the epigenetic level. Understanding how V(D)J recombination is regulated, in turn, provides insight into how misregulation of this process can lead to genomic instability and frank oncogenesis.

Figures and Legends

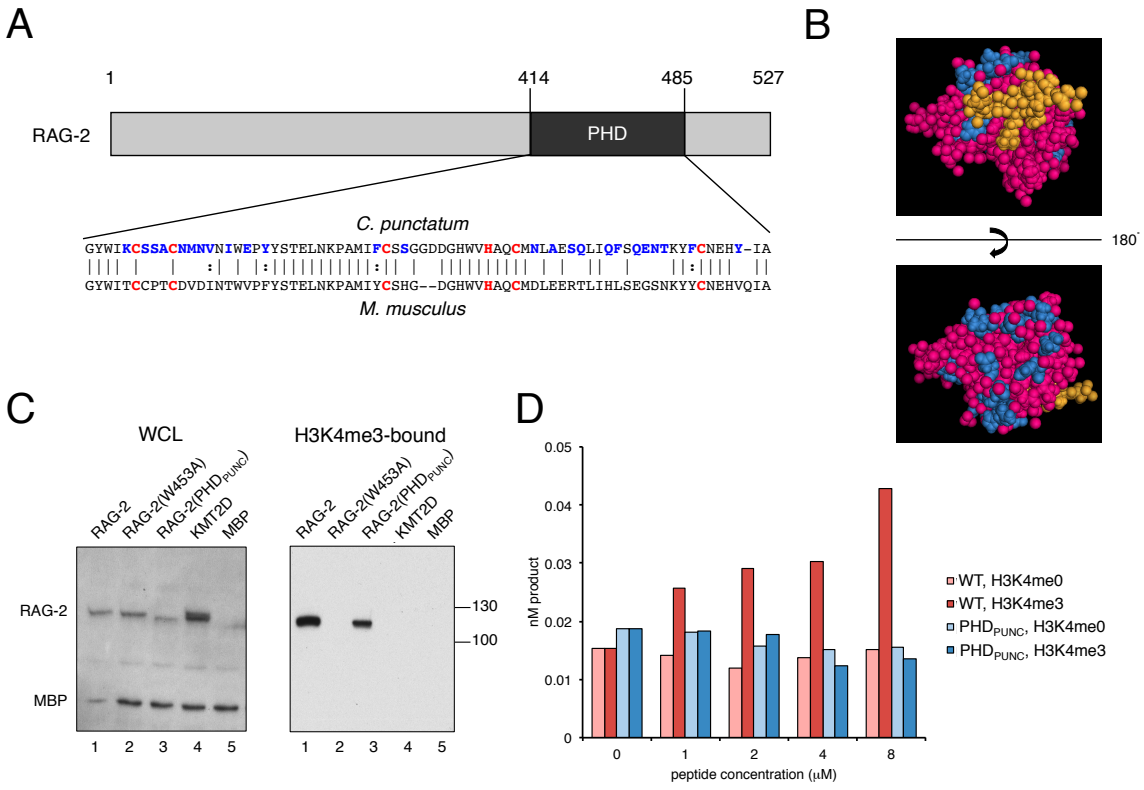


Fig. 1

Figure 1. Association of RAG-2(PHDPUNC) with H3K4me3 in the absence of allosteric activation. *A*, *above*, representation of *M. musculus* RAG-2 highlighting the plant homeodomain (PHD) sequence in the interval from residue 414 through 485. *Below*, alignment of PHD sequences from *C. punctatum* and *M. musculus*. Residues invariant among all PHD sequences are marked in red; residues in *C. punctatum* that differ from those of the mouse are marked in blue. Conserved residues are indicated by vertical lines; similar residues are indicated by vertical dots. *B*, *above*, space filling model of *M. musculus* RAG-2-PHD finger in complex with H3K4me3 peptide (Protein Data Bank ID 2V89). H3K4me3 peptide is indicated in orange, residues in *M. Musculus* that differ from those of *C. punctatum* are indicated in blue, remaining residues are indicated in magenta. *Below*, the same structure, rotated 180˚downward. *C*, binding

of RAG-2 and the RAG-2(PHD_{PUNC}) chimera to H3K4me3 peptide. *Right*, whole cell lysates of HEK293T cells expressing full-length, wild-type RAG-2, RAG-2(W453A), RAG-2(PHD_{PUNC}) or a fragment of KMT2D, all as MBP fusions, were incubated with bead-bound H3K4me3 peptide. Bound protein was detected by immunoblotting. *Left*, a portion (10%) of each lysate was assayed in parallel. *D*, assay for coupled cleavage. MBP-cRAG-1CT and MBP-RAG-2 or MBP-RAG-2(PHD_{PUNC}) were copurified and assayed for coupled cleavage of a radiolabeled 12-RSS and an unlabeled 23-RSS in the presence of increasing amounts of H3K4me3 or H3K4me0 peptide. Accumulation of nicked product at 1 hr is plotted against the concentration of H3K4me0 (pink and light blue) or H3K4me3 (red and dark blue). Wild-type RAG-2 is indicated in pink and red; RAG-2(PHD_{PUNC}) is indicated in light and dark blue.

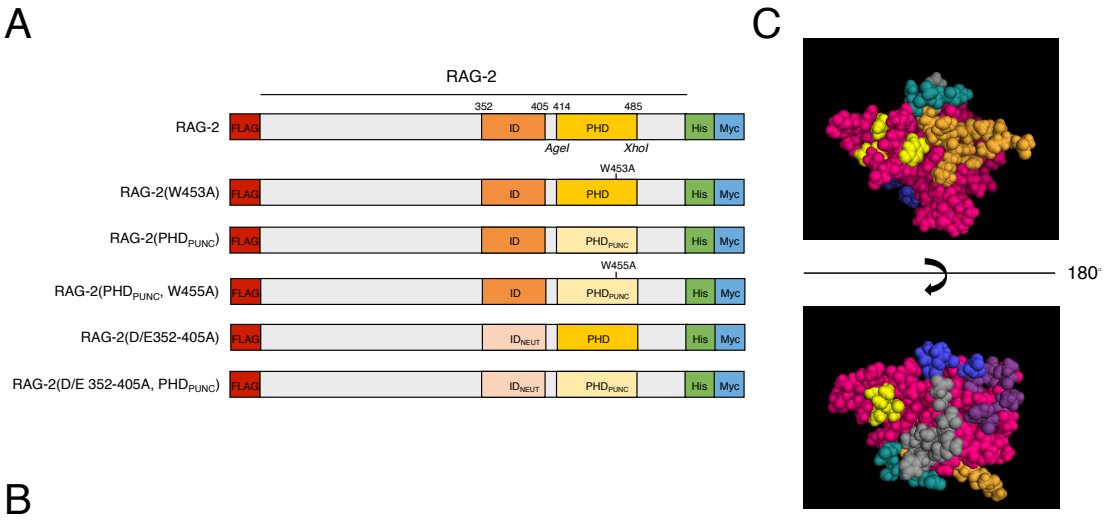


Fig. 2

Figure 2. Strategy for phylogenetic substitution and back mutation of the RAG-2 PHD. *A,* RAG-2 variants used in this study. Variants are identified at left. FLAG epitope (FLAG) as well as polyhistidine and c-Myc epitopes (His, Myc) are indicated. ID, inhibitory domain. PHD, plant homeodomain. PHD_{PUNC}, *C. punctatum* PHD finger; ID_{NEUT}, inhibitory domain bearing neutralizing mutations at acidic residues. Amino acid residues at domain boundaries are indicated above and numbered according to the mouse sequence. Positions of point mutations introduced into mouse and shark PHD fingers are indicated. *AgeI* and *XhoI* restriction sites, introduced to facilitate exchange of PHD cassettes, are shown. *B,* alignment of *M. Musculus* and *C. Punctatum* PHD sequences. Numbering follows the mouse sequence. Residues that are invariant among all PHD sequences are indicated in red; residues in *C. Punctatum* that differ from those in the mouse are indicated in light blue. Clustered mutations that convert residues in *C. punctatum* to the corresponding residues in the mouse PHD are indicated as follows: Group1,

dark blue; Group 2, dark purple; Group 3, grey; Group 4, teal; and Group 5, yellow. *C*, mapping of each mutation group onto a space-filling model of the mouse RAG-2 PHD finger in complex with H3K4me3. Groups are color-coded as in *B*. The H3K4me3 peptide is indicated in orange and remaining PHD residues are in magenta.

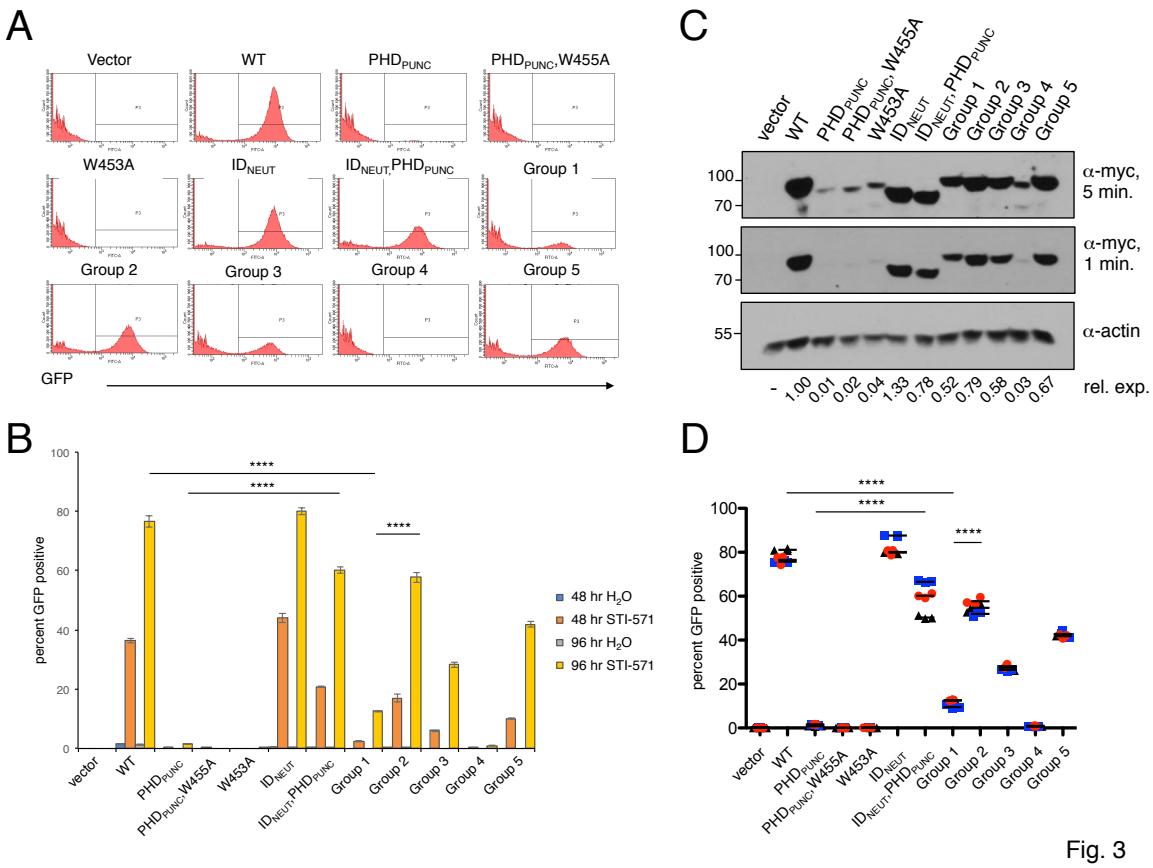


Fig. 3

Figure 3. RAG-2(PHDPUNC, Group1) exhibits normal accumulation but impaired V(D)J recombination activity. *A*, fluorometric assay for V(D)J recombination. RAG-2 variants were transduced into R2K3 cells, selected in puromycin for viral transductants and recombination was induced with STI-571; GFP-positive cells were detected by flow cytometry 96 hr after induction. ID_{NEUT}, RAG-2(D/E352-405A); ID_{NEUT}, PHD_{PUNC}, RAG-2(D/E352-405A, PHD_{PUNC}); Group 1 through Group 5, RAG-2(PHD_{PUNC}) variants bearing clustered mutations as defined in Fig. 2*B*. *B*, Quantified fluorometric assay. GFP-positive cells were counted at 48 and 96 hr after induction. Mean (n = 3) and standard deviation are indicated. Control samples were mock-induced with water (H_2O) as indicated at *right*. Significance was assessed for 96 hr samples by one-way ANOVA and Tukey's post-hoc test (****, $p < 0.0001$). *C*, *above*, anti-myc immunoblotting of

RAG-2 variants in R2K3 lysates at 96 hr after induction. Images at 5 min and 1 min exposure times are displayed. *Below*, immunoblotting of actin on the same membrane as *above*. Relative expression (rel. exp.) was determined by normalizing each anti-myc band in the 1 min exposure to the corresponding actin band, followed by normalization to wild-type. *D*, the RAG-2 variants assayed in *A* and *B* were transduced independently in triplicate and assayed in triplicate for recombination. The means of the technical triplicates for each biological replicate are plotted. Significance was assessed by one-way ANOVA and Tukey's post-hoc test (****, $p < 0.0001$).

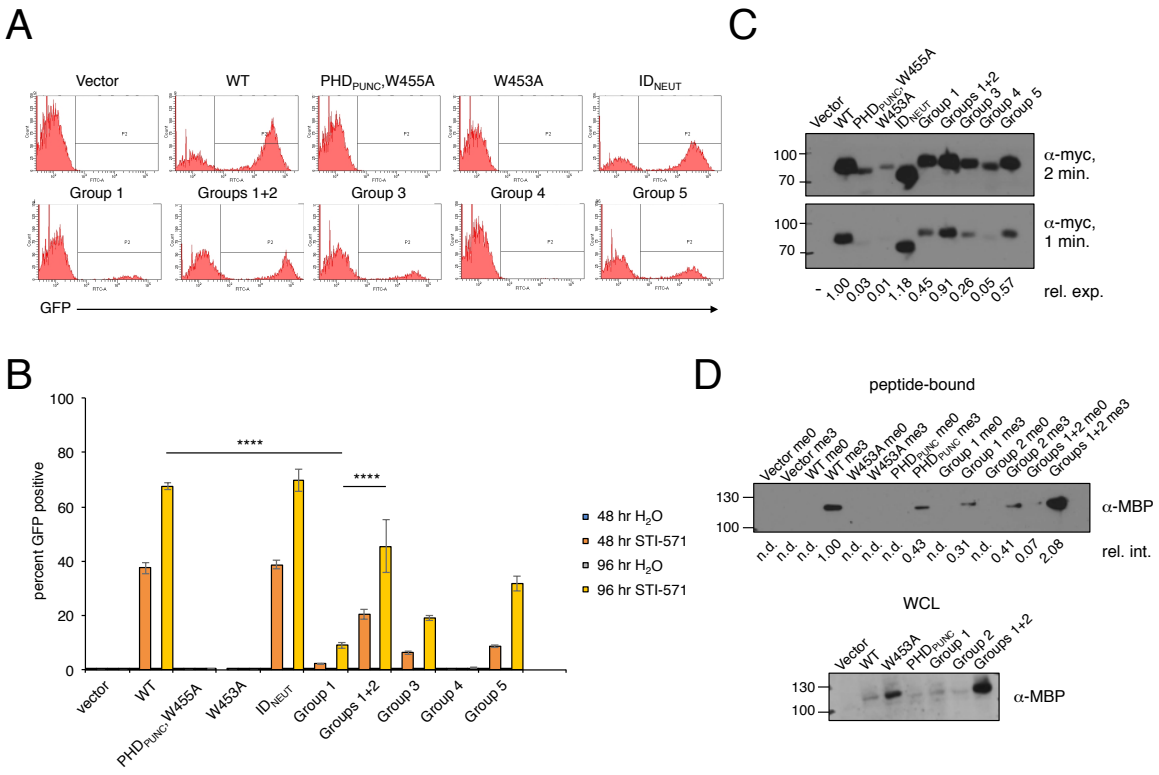
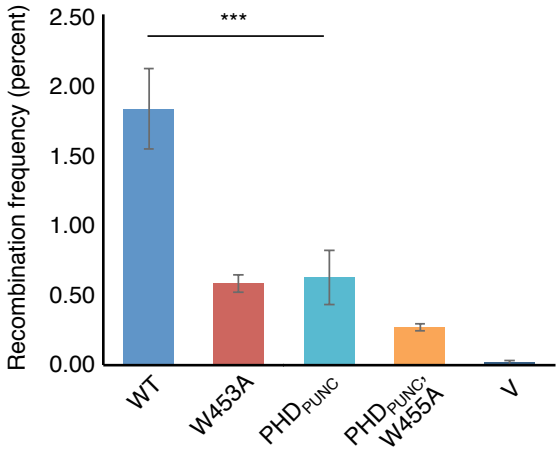


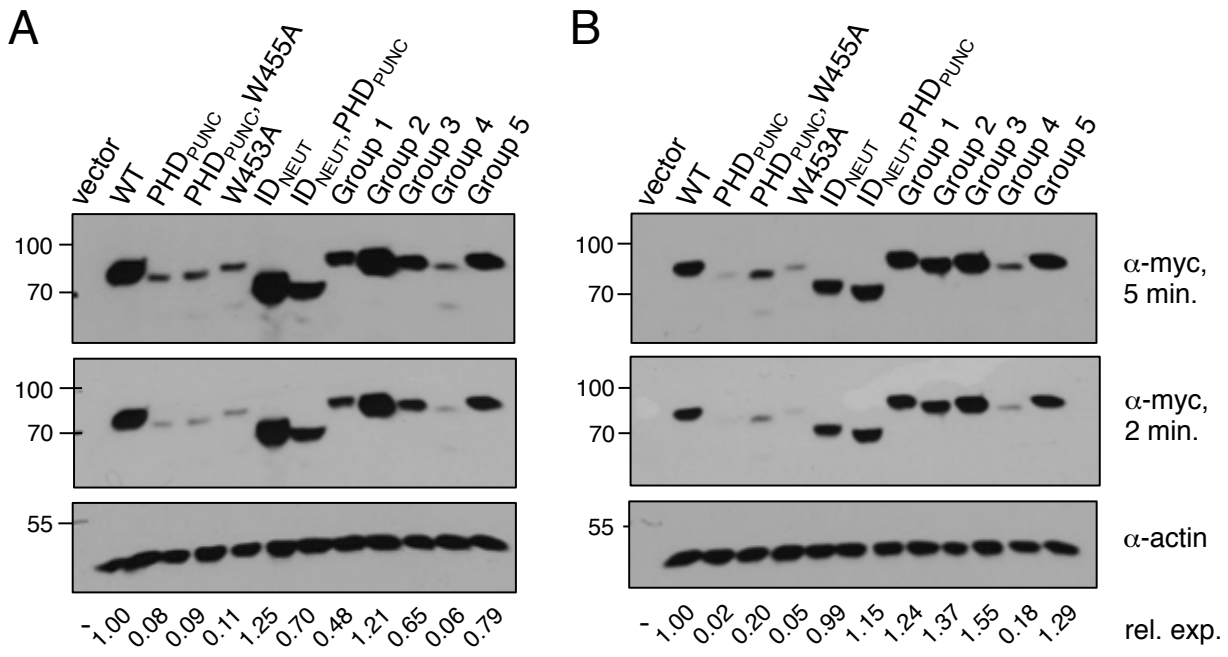
Fig. 4

Figure 4. The Group 2 mutation confers recombination activity on RAG-2(PHDPUNC, Group1). *A*, RAG-2 variants as indicated were assayed for V(D)J recombination in the R2K3 fluorometric assay. Displayed are representative flow cytometric data at 96 hr after induction with STI-571. *B*, Flow cytometric data were quantified at 48 and 96 hr after induction. RAG-2 variants are indicated *below*. Control samples were mock-induced with water (H_2O) as indicated at *right*. Mean (n = 3) and standard deviation are indicated. Significance was assessed by one- way ANOVA and Tukey's post-hoc test (****, $p < 0.0001$). *C*, anti-myc immunoblotting of RAG-2 variants in R2K3 lysates at 96 hr after induction. Images at 2 min and 1 min exposure times are displayed. Relative expression (rel. exp.) was determined by normalizing each anti-myc band in the 1 min exposure to the corresponding actin band, followed by normalization to wild-type. *D, above*, assay for binding to H3K4me3. Lysates of HEK293T cells expressing RAG-2 variants as

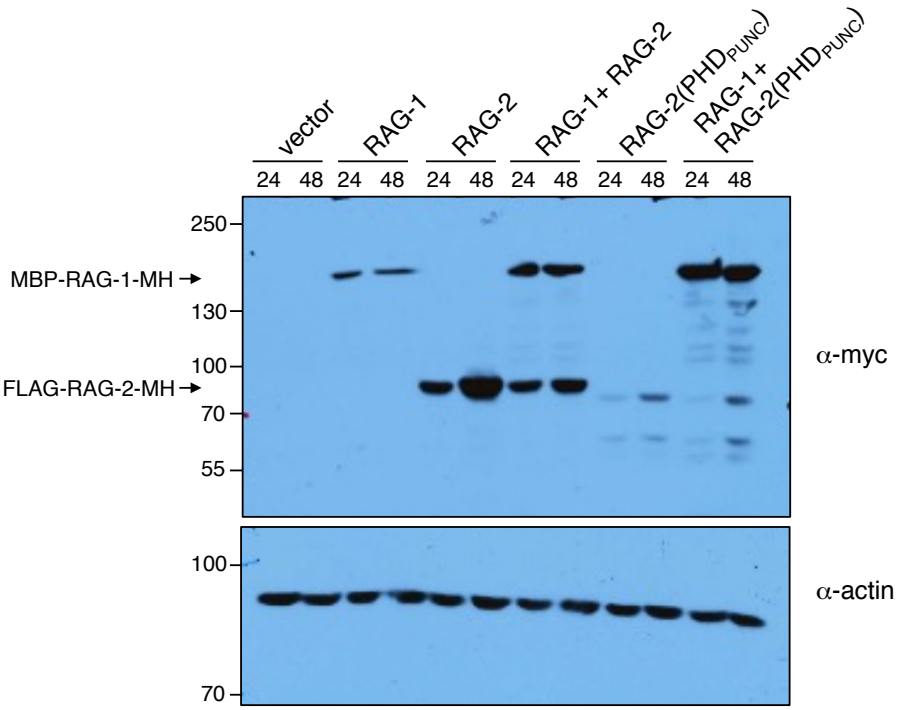
fusions to MBP (300 μg) were incubated with bead-bound H3K4me0 or H3K4me3 peptide. Bound protein was detected by immunoblotting for MBP. The relative intensity (rel. int.) of each band, in comparison to wild-type, is indicated; n.d., not detected. *Below*, 50 μl of each whole-cell lysate was immunoblotted for MBP.



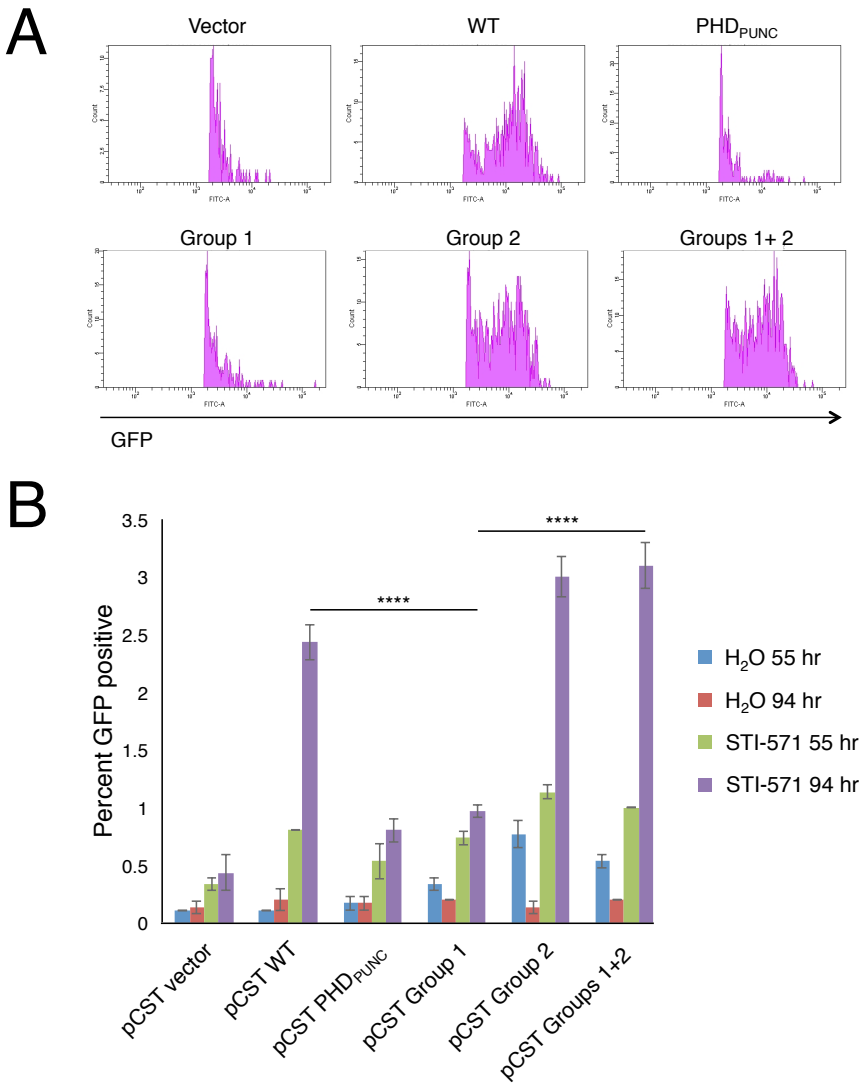
Supplemental Figure 1. Extrachromosomal assay for V(D)J recombination. NIH3T3 cells were transfected with the extrachromosomal substrate pJH200, MBP-tagged, full-length RAG-1 and MBP-RAG-2-myc-his variants as defined below. V, transfection with vector in place of RAG-2. V(D)J recombination in eukaryotic cells removes a prokaryotic transcriptional stop from pJH200, permitting subsequent expression of chloramphenicol acetyltransferase in *E. coli*. Plasmid DNA was recovered at 48 hr and transformed into *E. coli*. Transformants were scored for chloramphenicol and ampicillin resistance. The frequency of recombination (chloramphenicol + ampicillin resistant/ampicillin resistant) is plotted for each RAG-2 variant.



Supplemental Figure 2. Accumulation of RAG-2(PHDPUNC) mutant proteins in vivo. Three independent biological replicates, involving independent viral infections, were conducted for the R2K3 rearrangement assay. An immunoblot assay corresponding to the first replicate is displayed in Fig. 3C; immunoblot assays corresponding to the additional replicates are displayed here. A, *above*, anti-myc immunoblotting of RAG-2 variants in R2K3 lysates at 96 hr after STI-571 induction; 5 min and 2 min exposure times are displayed. Immunoblots correspond to the recombination assays indicated in blue in Fig 3D. *Below*, immunoblotting for actin on the same membrane as above. Relative expression (rel. exp.) was determined by scanning anti-myc bands at 2 min exposure time. For each band the area under the curve (AUC) was first normalized to the corresponding AUC for actin and then to the AUC for wild-type RAG-2. B, Immunoblots corresponding to the recombination assays indicated in black in Fig 3D. Data are presented as in panel A.



Supplemental Figure 3. Overexpression of RAG-1 fails to enhance accumulation of RAG-2(PHDPUNC). RAG-2 constructs were expressed from the pCLIP2a vector while RAG-1 constructs were expressed from the pcDNA3.1 vector. HEK-293T cells were co-transfected with: empty vectors; RAG-1 and pCLIP2a; RAG-2 and pcDNA3.1, RAG-1 and RAG-2; RAG-2 (PHD_{PUNC}) and pcDNA3.1; or RAG-1 and RAG-2(PHD_{PUNC}). RAG-1 and RAG-2 variants were detected by immunoblotting using an anti-Myc antibody at 24 and 48 hr after transfection. To control for loading -actin was used.



Supplemental Figure 4. The clustered Group 2 back mutation rescues V(D)J recombination activity of RAG-2(PHDPUNC, Group 1). *A*, Fluorometric assay for V(D)J recombination in R2K3 cells transduced with RAG-2 variants as described in Fig. 3. Variants were transduced using the pCST lentiviral vector, which encodes a Thy1.1 selection marker. Transduced cells were enriched by magnetic bead-based selection for Thy1.1, arrested with STI-571 and assayed for inversional V(D)J recombination by flow cytometry at 55 and 94 hr after induction. Representative flow cytometric results at 94 hr, gated on GFP+ cells, are displayed. *B*,

Quantified fluorometric assay for V(D)J recombination. Percent GFP-positive cells was determined at 55 and 94 hr after induction. Mean (n = 3) and standard deviation are indicated. Significance was assessed for 94 hr samples by one-way ANOVA and Tukey's post-hoc test (****, $p < 0.0001$).

Table S1

Primer Name	**Sequence 5' to 3'**	**Purpose**
pCST 5'	AGCCCTCACTCCTTCTCTAG	Sequencing pCST insert
pCST 3'	CGGGCCAGTGAATTGTAATACGACTCACTATAGGGAGGCGG	Sequencing pCST insert
pCLIP 5' for 2019	CAGCTCCTGGGCAACGTG	Sequencing pCLIP insert
pCLIP 3'rev 2019	CCTAGATGCATGCTCGACCTG	Sequencing pCLIP insert
Cla1 FLAG RAG2 2019	GAGAatcgatGACTACAAAGACGATGACGACAAGCttTCCCTGCAGATGGTAACAGTGGG	PCR, to introduce FLAG, remove RAG2 from pcDNA3.1 vector to put into pCLIP2a vector
Mlu1 linker RAG2 rev 2019	gagaACGCGTGATGGTGATGGTGGAGGTGATGGTGGAATTCATCAAACAGTCTTCTAAGGAAGGATTTCTTGG	PCR, to remove RAG2 from pcDNA3.1 vector and put into pCLIP2a vector

Supplemental Table 1 : PCR primers

Experimental Methods

Coupled Cleavage Assays. Assays for coupled cleavage of double-stranded oligonucleotide substrates HL-44/45 (12-RSS) and HL-46/47 (23-RSS) were performed as described (20), except that reactions contained 100nM of RAG tetramer and were preincubated for 30 min in cleavage buffer before addition of $MgCl_2$.

Assay for endogenous recombination of immunoglobulin gene segments. The retroviral vector pCLIP2a (152) was programmed to express puromycin N-acetyl transferase and FLAG-RAG2-6His-Myc variants. Viral particles were generated by cotransfection of pCLIP2A constructs and pCL-ECO packaging vector into HEK293T cells and concentrated by centrifugation. R2K3 cells were infected by spin inoculation in the presence of 10 μg/mL polybrene. Cells were maintained under selection with 0.8 μg/ml puromycin for 7-9 days. Recombination was induced in R2K3 cells (10 ml at 10^6 cells/ml) by addition of STI-571 to a final concentration of 3 μM for 48 or 96 hr.

Cell Culture. Cell lines were propagated at 37° in 5% CO_2. HEK293T cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 1X penicillin-streptomycin-glutamine (PSG). R2K3 cells (153)were grown in RPMI1640 medium supplemented with 10% FBS, 50 μM 2-mercaptoethanol, 1X PSG, 0.7X MEM non-essential amino acids, 1 mM sodium pyruvate and 10 mM HEPES. R2K3 cells containing an integrated PMX-INV recombination substrate (153) were provided by Dr. Barry Sleckman (Weill Cornell Medicine)

Antibodies. Commercial antibodies against the following proteins were used in this study: actin (clone AC-40, Sigma Aldrich, Cat #A3853); c-myc (clone 9E10, Santa Cruz Biotechnology, Cat#sc-40).; MBP (clone N-17, Santa Cruz Biotechnology, Cat#sc-809); histone H3K4me3 (clone AW304, Millipore Cat#05-791R).

Immunoblotting. Immunoblotting was performed in PBST (1X PBS, 0.1% Tween) containing 5% nonfat dry milk. For immunoblotting, 60 μg of protein was fractionated by 8% SDS-PAGE and transferred to a 0.45um PVDF-plus membrane. Membranes were blocked for 1 hr at 37˚C with 5% nonfat dry milk in PBST, incubated with primary antibody for 1 hr at room temperature or overnight at 4˚C, washed 3 x 15 min in PBST, incubated with secondary antibody for 1 hr at room temperature and washed 3 x 15 minutes in PBST. Secondary antibody was detected by enhanced chemiluminescence (Amersham ECL Prime).

Expression constructs. DNA fragments encoding full-length wild-type RAG-2 and RAG-2(W453A), fused in-frame to a cassette encoding an amino-terminal maltose-binding protein (MBP) tag followed by a tobacco etch virus (TEV) protease cleavage site (MBP-TEV), were cloned between the BamHI and ApaI restriction sites of pcDNA3.1 to produce the plasmids pcDNA3.1-MBP-TEV-wtRAG2 and pcDNA3.1-MBP-TEV-RAG2(W453A) (20). To facilitate substitution of variant PHD coding sequences for the mouse wild-type PHD coding sequence, silent C1242T, A1455T and A1456C mutations were placed within the RAG-2 coding sequence of pcDNA3.1-MBP-TEV-wtRAG2 so as to introduce AgeI and XhoI restriction sites at the borders of the PHD coding interval. The resulting plasmid, pcDNA3.1-MBP-TEV-wtRAG2-AX, after cleavage with AgeI and XhoI and removal of the wild-type PHD coding sequence, was then used as a recipient for synthetic duplex DNA fragments encoding PHD variants ordered as gblocks from IDT.

The retroviral vector pCLIP2a (54,152) was used for transduction of RAG-2 variants into recipient cells. Sequences encoding wild-type RAG-2 or RAG-2 variants were amplified from pcDNA3.1-based vectors by PCR using primers Cla1FLAGRAG2for2019 and Mlu1linkerRAG2rev2019. The resulting amplicons were cloned between the ClaI and MluI restriction sites of pCLIP2a to produce viral vectors encoding RAG-2 variants flanked at the amino terminus by a FLAG epitope and at the carboxyl terminus by a His_6-Myc_3 tag. The retroviral vector pCST(160,161), encoding the selectable marker Thy1.1, was used in some experiments as specified for transduction of RAG-2 variants containing an amino terminal FLAG epitope.

The nucleotide sequences of oligonucleotide primers and synthetic duplex DNA fragments used in the construction of expression vectors are provided in **Supplemental Table 1**.

Purification of RAG protein. MBP fusions to full-length RAG-2 variants and an MBP fusion to a fragment of RAG-1 spanning amino acid residues 383-1040 (cRAG1-ct) were co-expressed in HEK293T cells and purified by amylose affinity chromatography as described (20). Protein concentration was determined by 8% SDS-PAGE and Coomassie staining, followed by quantification using ImageJ.

H3K4me3 binding assay. Binding assays were performed as described (6), with modifications. Biotinylated peptides derived from histone H3, unmethylated or trimethylated at lysine 4, were purchased from Millipore. HEK293T cells, expressing RAG-2 variants, were lysed in a buffer containing 50 mM Tris (pH 7.5), 300 mM NaCl, 1 mM PMSF, 1% NP-40, 1% deoxycholic acid, 0.1% SDS, and a cocktail of protease inhibitors (Roche). After centrifugation at 16,000 g for 15 min, aliquots containing 300 μg protein were precleared with 50 μl of a 50% streptavidin agarose slurry (Novagen) that had been equilibrated with binding buffer (50 mM Tris [pH 7.5], 300 mM

NaCl, 1 mM PMSF and protease inhibitors). Precleared samples were then incubated with 5 μg peptide prebound to streptavidin agarose for 4 hr at 4°C. Beads were washed three times with binding buffer supplemented with 0.1% NP-40. Bead-bound protein was released by heating to 95˚C in SDS-PAGE buffer, fractionated by SDS-PAGE and detected by immunoblotting for MBP.

Assay for extrachromosomal recombination. Assays for recombination of extrachromosomal substrates pJH200 (26) were performed as described (17), with modifications. MBP-RAG-1-myc-his (10 μg), MBP-RAG-2-myc-his or RAG-2 variants as indicated in Supplemental Figure 1 (10 μg) and 4 μg pJH200 (Addgene plasmid #13327) were transfected into NIH3T3 cells using the TransIT-LT-1 transfection reagent (Mirus Bio, CAT# MIR2306). At 48 hr post-transfection, plasmid DNA was recovered by a modified Hirt extraction (Qiagen, CAT #27104). DNA (4 μl, about 40 μg) was transformed into 50 μl DH5α Max Efficiency cells (Thermo Fisher, Cat #18258012). A 5 μl aliquot of the transformation mixture was plated on LB agar containing 50 μg/ml ampicillin and 300 μl was plated on LB agar containing 50 μg/ml ampicillin and 20 μg/ml chloramphenicol. Plates containing ampicillin alone were scored after 16 hr at 37˚C, while plates containing ampicillin and chloramphenicol were scored at 20 hr.

Chapter 2: An antigen receptor locus-specific capture sequencing method reveals structural variations in a mouse model of mistimed V(D)J Recombination

Abstract

V(D)J recombination is catalyzed by the recombination-activating gene protein (RAG) recombinase, which consists of RAG-1 and RAG-2 subunits. RAG-induced DNA cleavage is constrained to G0/G1 of the cell cycle through the periodic destruction of RAG-2 protein. Phosphorylation at RAG-2 T490 leads to the subsequent ubiquitination and degradation of RAG-2 via the proteasome. The RAG-2 T490A mutation uncouples RAG-mediated DNA cleavage and V(D)J recombination from the cell cycle. Against a p53-null background the RAG-2 T490A mutation induces lymphoid malignancies characterized by complex chromosomal translocations involving antigen receptor loci. To further characterize the genomic rearrangements, present in lymphoid tumors from RAG-2(T490A), p53-deficient mice, we developed a capture-sequencing method to enrich for genomic rearrangements involving antigen receptor loci. This method is capable of identifying endogenous V(D)J rearrangements as well as translocations involving immunoglobulin and T cell receptor genes. On a global scale, these tumors are characterized by high levels of interchromosomal rearrangements.

Introduction

Antigen Receptor Loci Organization

There are seven antigen receptor loci that encode for B cell receptors (BCRs) and T cell receptors (TCRs), immunoglobulin heavy chain (IGH), immunoglobulin light chain lambda (IGL), immunoglobulin light chain kappa (IGK), TCRα/δ (TRAD), TCRβ (TRB), and TCRγ (TRG). All the loci contain some combination of variable(V), diversity (D), junctional (J) and constant (C) gene segments, but the loci differ in the number of gene segments available. Each locus has a distinct genomic organization and exhibits differential modes of regulation.

BCRs are comprised of heavy and light chains encoded by the IGH, IGK and IGL loci. The murine immunoglobulin heavy chain (IGH) genes are assembled by the joining of V_H, D_H, and J_H gene segments that span 2.5Mb of the genome on chromosome 12 (162,163). The IGH locus contains 150 V_H segments, 10-13 D_H segments, 4 J_H segments, and 8 C_H gene segments. About 100kb separates the V_H segments from the D_H gene cluster. The D_H cluster is flanked on the 5' end by gene segment DFL16.1 and on the 3' end by DQ52. DQ52 is less than 1kb upstream of the J_H cluster. Downstream of the J_H cluster lies the Eμ intronic enhancer and the exons encoding Cμ, Cδ, Cγ3, Cγ1, Cγ2b, Cγ2a, Cε, and Cα. Further downstream of Cα is the 3' regulatory region(162). In contrast to IgH, the IGK and IGL loci contain only V and J gene segments. For example, IGK is on mouse chromosome 6 and contains 96 function V_K segments, which span 3.2Mb, 4 J_K segments, and $C_{K(163)}$.

TCRs are comprised of either α and β chains or γ and δ chains. The TRG is the smallest TCR locus, spanning less than 200kb, while the TRAD locus spans over 1Mb(164). TRG has four copies of the V-J-C gene segment cassette. The TRAD locus is comprised of the TCR (TRD) locus embedded within the TCR (TRA) locus on chromosome 14. This locus contains 104 V segments, which can rearrange to two Dδ and Jδ segments or to 61 Jα segments (163). The TRB locus spans 700kb on chromosome 6 and contains 35 V segments upstream of a duplicated cluster, each containing 1 D segment, 7 J segments, and a set of C exons. These clusters are termed TRβ1 and TRβ2.

B cell development

B cells originate in the bone marrow from lymphoid progenitors derived from hematopoietic stem cells, mature in the bone marrow, and eventually move to the spleen and secondary lymphoid tissues. The IGH locus is the first to undergo rearrangement. These gene segments are rearranged in an ordered fashion beginning with D-to-J_H rearrangements, followed by V_H-to-DJ_H rearrangements until a productive rearrangement is created (113). B cells transition from early pro-B cells which undergo D-to-J_H rearrangement, to late pro-B cells which undergo V_H-to-DJ_H rearrangement, to large pre-B cells that contain a completed V(D)J_H rearrangement, to small-preB cells which undergo V_L-to-J_L rearrangement, to immature B cells containing a fully rearranged heavy and light chains. At the pro-B and pre-B stages, B cells express a surrogate light chain which associates with the heavy chain forming a pre-B cell receptor complex on the surface of the cell until the IGK or IGL light chain locus is productively rearranged(165,166). The pre-BCR induces proliferative expansion of pre-B cells, downregulation of the surrogate light chain, and downregulation of RAG-1 and RAG-2 transcription(166). Upon successful light

chain gene rearrangement, a naive B cell is formed and is ready to encounter antigen. While in the bone marrow, B cells undergo positive selection to select for cells that can recognize ligand and signal properly. B cells then proceed through negative selection in which self-reactive B cells either are clonally deleted through apoptosis(167-169) or undergo another round of V-to-J light chain gene recombination, termed receptor editing, to eliminate self-reactivity(170). After selection, B cells then migrate to the spleen and differentiate into mature naïve B cells.

Cells must encounter antigen in order to become activated. These interactions with antigen occur predominantly in secondary lymphoid tissues such as the spleen, lymph nodes, and Peyer's patches. These tissues are specialized to filter body fluids, to capture and display foreign antigens for B cell recognition, and to recruit lymphocytes from the blood. Within a lymphoid follicle, follicular dendritic cells can capture and display opsonized antigens. These antigens can be displayed to B cells and activate B cell receptors (BCRs). Signaling through the BCR, with the help of co-receptor activation, T cell help, and/or cytokine signals initiates B cell activation(171). Activation can be T cell dependent or independent. The BCR and bound antigen is then internalized, antigen is enzymatically processed, and antigenic peptides are displayed on MHC class II for presentation to T cells.

After activation depending on the signals, B cells can differentiate to form germinal center B cells, plasma cells, memory B cells, or mature B cells(171). B cells that become germinal center B cells can undergo affinity maturation, somatic hypermutation by AID, and selection within the germinal center. B cells that are differentiating into plasma cells and still actively proliferating are called plasmablasts, but upon the cessation of proliferation they become plasma

cells/antibody secreting cells. Memory B cells can be reactivated by subsequent antigen exposures and are poised to rapidly differentiate. Upon re-exposure to antigen, memory B cells may undergo class switch recombination to change the isotype of the BCR from IgM and IgD to IgG, IgA, or IgE (171).

T cell development

T cells develop from lymphoid progenitors which originate from bone marrow progenitors and then migrate to the thymus where they undergo maturation, selection, and export to peripheral tissues. T cells begin as $CD4^-CD8^-$ double negative (DN), become $CD4^+CD8^+$ double positive (DP), and then finally become $CD4^+ CD8^-$ or $CD4^-CD8^+$ single positive (SP). Within the thymus, T cells undergo T cell receptor (TCR) rearrangement to eventually generate $CD4^+CD8^+$ double positive thymocytes. TCRs are encoded within the TRAD, TRB, and TRG loci. Recombination of the TRG, TRB, and TRD loci all occur during the double negative stages of thymocyte development. The two possible types of TCR are γδ TCR and αβ TCRs. Productive rearrangement of TRG and TRD results in expression of a γδ TCR, whereas successful rearrangement of TRB leads to a pre-TCR comprised of TCRβ with a pre-TCRα (172). The pre-TCR can signal to downregulate RAG expression, increase proliferation, mediate β chain allelic exclusion, and differentiation toward the DP thymocyte stage. At the DP stage RAG is re-expressed and TRA rearrangement begins.

Each locus is controlled and recombined differently. TRG rearrangement is dependent on IL-7 signaling (172). IL-7 activates STAT5 which can bind to Jγ promoters and 3'Eγ enhancer. Recombination at TRB is developmentally controlled, with Dβ-to-Jβ recombination preceding

Vβ-to-DβJβ recombination. This control is dependent on the Eβ enhancer and the nature of the RSSs(172). The Vβ-to-DJβ rearrangement is subject to allelic exclusion, so that T cells only express one function TRB gene. This is enforced through feedback inhibition established by pre-TCR signaling. The TRAD locus consists of the TRD locus embedded within the TRA locus. Vδ-toDδ-toJδ recombination occurs in DN T cells, whereas Vα-to-Jα recombination occurs in DP T cells(172). This developmental control is enforced through the differential activation of developmental stage specific enhancers.

Within the thymus, DP thymocytes undergo positive selection based on the ability to interact with MHC-I or MHC-II. This ensures the T cells have the ability to recognize MHC molecules. Cells that can properly respond to MHC-I or MHC-II receive survival signals while those that do not interact well will die. Next a thymocyte undergoes negative selection, or clonal deletion, to remove self-reactive clones. Thymocytes that interact strongly with self-antigens die by apoptosis, but some survive as regulatory T cells (Tregs). Tregs are beyond the scope of this dissertation and will not be covered in detail. Cells that survive positive and negative selection then exit the thymus as mature naïve T cells.

Naïve T cells encounter antigens and costimulatory ligands presented by dendritic cells. This activation requires engagement of the TCR and a costimulatory molecule by MHC-II and MHC-I for CD4+ and CD8+ T cells respectively, and results in IL-2 production, proliferation, and differentiation into effector cells that migrate to various sites and engage in pathogen clearance. The majority of effector T cells apoptose following infection, but a subset of T cells persists as long-term memory cells that can protect against a subsequent infection. Beyond activation, T

cells can further differentiate into various types of memory T cells such as stem cell memory, central memory, and effector memory T cells. These developmental stages are beyond the scope of this dissertation and will not be explored further.

Temporal regulation of V(D)J recombination

V(D)J recombination is restricted to the G0/G1 phase of the cell cycle through the periodic degradation of RAG-2(61). RAG-2 is phosphorylated at T490 by CyclinA/CDK2 at the G1-to-S transition leading to subsequent ubiquitylation by the Skp2-SCF ubiquitin ligase, followed by proteasomal degradation(57-60). By limiting the formation of RAG-mediated DNA DSBs to G0/G1, the resulting DNA ends are sequestered from homologous recombination and repaired by NHEJ which predominates during G0/G1. This coupling of V(D)J recombination to the cell cycle suppresses genomic instability and lymphoid tumorigenesis(62). Mutation of the T490 phosphorylation site to alanine permits RAG-2 expression throughout the cell cycle(62). This constitutive RAG expression leads to accumulation of DNA DSB throughout the cell cycle and exposes DNA ends to both NHEJ and homologous recombination machinery. The mouse model employed in this dissertation contains the T490A mutation crossed into a p53 deficient background. These mice exhibit complex chromosomal translocations involving the antigen receptor loci(62).

Spatial regulation of V(D)J recombination

Three-dimensional genome organization and spatial proximity can influence genome wide chromosomal rearrangements and translocations(173). Through the use of genome-wide chromatin conformation capture using proximity ligation (Hi-C), mouse chromosomes with close

spatial proximity in the nucleus were found to correlate with increased participation in translocations(173). Additionally, spatial organization appears to be similar between species(173). For example, the frequency of IgH, IgK or IgG to c-myc translocations in Burkitt's lymphoma is positively correlated with the reciprocal nuclear distance between each loci(174).

In addition to nuclear proximity, synapsis of antigen receptor gene segments is required for cleavage by RAG. These segments span large regions of the genome. In the RSS capture model, RAG binds one RSS and then captures a second RSS to form the synaptic complex. V-to-DJ recombination requires conformational changes within the antigen receptor locus to reduce the distance between the V gene segments and the DJ segment. Chromatin looping can bring together distant gene segments. Chromatin looping mediated by the CCCTC binding factor (CTCF) has been implicated in the regulation of V(D)J recombination(175). CTCF is a transcription factor containing 11 zinc-finger domains that bind to many sites in the mammalian genome and is enriched at antigen receptor loci(118). The murine IgH locus spans 2.7Mb, so that locus contraction and looping are needed to aid in rearrangement. There are 144 putative CTCF-binding sites within the IgH locus which appear to be evolutionarily conserved and exhibit a strand-orientation bias with respect to V_H gene transcription(176).

Through the introduction of bait RAG-generated breaks, it was found that RAG "off-target" breaks, occurring at the CAC motif of cryptic RSSs, are largely confined to convergent, CTCF-binding element (CBE)-flanked domains containing the bait RSS pair(177). This observation is consistent with linear tracking of RAG along chromatin within a loop(177). V_H-associated CBEs appear to also play a role in V_H gene segment utilization(178).

CTCF-mediated looping can also regulate RSS availability, enhancer activation, and directional tracking of RAG along chromatin(118). Experiments involving introduction of ectopic CBEs at the TRAD locus demonstrated that CTCF is able to regulate V(D)J recombination by preventing RSS synapsis through the separation of RSSs into discrete loop domains(118). This effect was independent of transcription and RSS accessibility (118).

A key regulatory region, the intergenic control region 1 (IGCR1) resides between the V_H and D_H clusters of the IgH locus. IGCR1 plays a role in B cell development, suppresses activation of germline V_H segments, balances proximal versus distal V_H segment use, and is required for lineage specific V_H-to-DJ_H recombination (179). These functions of IGCR1 are dependent on CTCF-binding elements CBE1 and CBE2, although individual deletions of CBE1 or CBE2 produce only slight impairment of function, suggesting cooperative activity(180). Higher order looping can allow gene loci to migrate further away from their respective chromosome territory within the nucleus. Data revealing looping of the TRA locus suggests that higher order DNA looping can help direct RAG to specific loci(181,182).

Oncogenic genomic aberrations and V(D)J recombination

Many human cancers, including leukemias and lymphomas, are characterized by genomic rearrangements such as translocations. Oftentimes, these translocations are insufficient on their own to cause malignancies but serve as initiating events, after which secondary mutations can drive development of cancer. For example, in acute lymphoblastic leukemia (ALL) 25% of cases contain the ETV6-RUNX1 fusion protein(56). Additional somatic mutations are required for

development of overt leukemia. RAG-mediated recombination is the dominant mutational process that promotes development of ETV6-RUNX1-positive pre-B-ALL(56).

Genomic rearrangements are initiated by DNA double stranded breaks (DSBs) which can arise physiologically or by genotoxic insult. To generate a translocation, two DSB must be present simultaneously at the two participating loci, the loci must be in proximity or move into proximity, and DNA DSB repair must be available to join the broken DNA into a translocation(183). The two major sources of physiological DSBs in B cells are V(D)J recombination and class switch recombination. Improper repair of these physiological DSBs can lead to genomic rearrangements such as translocations. The origins of translocations can be attributed to recurrent DNA damage, frequent nuclear interactions, and random rearrangements(184). Additionally, the mode of DSB repair can also directly contribute to genomic abnormalities. In particular, lymphoid malignancies are often characterized by recurrent chromosomal rearrangements involving the activation of oncogenes(183). This commonly results from the generation of an oncogenic fusion protein or by deregulating oncogene expression by linking it to strong transcriptional elements of an antigen receptor locus(30,183,185).

Aberrant V(D)J recombination can be carcinogenic either through creation of off-target DSB or through the creation of DNA by-products. In addition to the coding joint formation, V(D)J recombination creates a signal joint comprised of the two RSSs and the intervening DNA fragment. Initial work demonstrated that RAG is capable of acting on these signal joints in vitro and subsequent studies showed RAG can support their transposition in the genome(186,187). Investigation of the ability of RAG to perform intermolecular recombination between two

plasmids led to proposal of a release and capture model in which coding ends are released from the post-cleavage complex and then recaptured by RAG(188). In vitro, RAG can cleave plasmid substrates both in cis and in trans. V(D)J recombination in trans has potential to lead to chromosomal rearrangements and genomic instability. One line of evidence implicating V(D)J recombination in genome instability came from detection of V(D)J recombination by-products in mouse thymocytes and their ability to be reintegrated into the genome(187,189). RAG exhibits a preference for GC rich DNA at the site of transposition(190). Episome-based assays demonstrated that RAG proteins can perform transposition in vivo, but at a low frequency(191). Interestingly, this work also revealed aberrant events in which the signal end fragment was integrated in the genome and accompanied by large scale deletions or translocations(191). Modes of regulation restricting these transposition events were discussed in Chapter 1 under the evolution of RAG-1 and RAG-2.

Three major types of translocations have been identified. The first is termed a type I translocation and involves illegitimate V(D)J recombination between a legitimate Ig/TCR gene segment and an illegitimate proto-oncogene locus. These translocations result from mis-targeting of RAG. Type II translocations involve legitimate Ig/TCR gene segment breaks, and DSB within a proto-oncogene generated by another mechanism. Type II translocations thus involve three DSBs, two at Ig/TCR segments and one at a proto-oncogene and result from mistakes in repair of the V(D)J intermediates. Type III translocations involve both targeting and repair mistakes by RAG and the repair machinery(192).

Off target V(D)J Recombination

V(D)J recombination can catalyze both DNA transposition and translocation in vitro and in vivo. Many lymphoid malignancies are characterized by chromosomal translocations involving antigen receptor genes and proto-oncogene loci, suggesting that these rearrangements could be RAG mediated(56). Many of these translocations can be attributed to off target RAG activity on genomic DNA containing cryptic RSS-like elements(193). Cryptic RSSs resemble an RSS, and are estimated to be present once every 600bp in the genome(194). Rearrangements involving cRSS can lead to oncogenic translocations and deletions(183,195-197) A canonical RSS consist of a heptamer (5'-CACAGTG) and a nonamer (5'ACAAAAACC) separated by either a 12 or 23bp spacer. Heptamers are cleaved by RAG 5' to the CAC motif, whereas the nonamer serves as a binding site for RAG-1. Single base-pair changes to an RSS can affect recombination. For example, mutation of the perfectly conserved third position of the heptamer blocks cleavage(198).

The C-terminus of RAG-2 is important for restricting off-target recombination. Mice deficient in p53 but expressing core-RAG-2 develop thymic lymphomas and exhibit more off-target rearrangements than their control p53 deficient counterparts(199,200). These core-RAG-2 p53 deficient mice also exhibit translocations between TRAD and IgH loci. These phenotypes are similar to those observed in the lymphomas of ATM deficient mice, suggesting that similar to ATM deficiency, core RAG-2 destabilizes the RAG post-cleavage complex(200). Since these breaks are not programmed to occur within the same lineage or the same cell, it was of interest to determine the mechanism underlying this translocation type. Nuclear accessibility of the TRAD and IgH loci increases in the absence of DNA repair factor ATM or the absence of the RAG-2

non-core(201). Translocations involving LMO2, TAL1, Ttg-1, and Hox11 have been documented as examples of illegitimate V(D)J recombination at cRSS sites emphasizing the importance of understanding potential off target V(D)J recombination in order to maintain genome integrity(202).

One mechanism of protection against these off-target recombination events has been the evolutionary depletion of cRSSs, specifically heptamer sequences, near RAG-1 binding sites(147). In particular by utilizing a ligation-mediated transferred-end capture assay which detects insertions of RSS-ended DNA fragments into the genome, core-RAG-2 thymocytes show a higher frequency of off target events at cryptic RSSs than wildtype mice, implicating the non-core region of RAG-2 in maintaining the integrity of the genome(193).

Another mechanism of protection from genomic instability involves the heptamer sequence and the stability of the post-cleavage complex containing RAG and the signal ends. After cleavage, RAG remains bound to the signal ends more tightly than the coding ends(203). Foot printing analyses revealed that this binding is dependent upon interactions between RAG and the heptamer sequences of the RSSs(203). Mutations that destabilize the signal-end complex in vitro allowed coding and signal ends to be repaired by inappropriate DNA double strand break repair pathways, such as alternative NHEJ, which is explained in further detail in the next section. Non-consensus heptamer sequences can destabilize the post-cleavage complex and allow for aberrant repair of DNA ends, potentially leading to genomic instability(203).

Double strand break repair of V(D)J recombination cleavage products

In addition to off-target RAG activity, the repair of DNA DSBs can also lead to genomic rearrangements. Unrepaired breaks that persist throughout a cell's development can lead to rearrangements between loci that usually are not accessible at the same time. Beyond the failure of break repair, the actual repair process itself can cause genomic rearrangements. There are three main pathways for repairing DSBs in mammalian cells, classical NHEJ (cNHEJ), alternative NHEJ (altNHEJ), and homologous recombination (HR)(87). cNHEJ factors are implicated in maintaining genomic integrity at the genome scale and is not considered to be a major error prone process. Artemis is an cNHEJ factor with both endo and exonuclease activity. During V(D)J recombination, Artemis opens the hairpins at coding ends. In Artemis-deficient murine embryonic stem cells, V(D)J coding joins lack deletions and display large palindromic insertions suggesting a hairpin opening defect(204). They also exhibit increased chromosomal instability which includes telomeric fusions. Deficiencies in cNHEJ factors on a p53-deficient background promote development of pro-B cell lymphomas characterized by oncogenic chromosomal translocations involving the IgH and c-Myc loci (85). These translocations exhibit microhomologies at the breakpoints suggestive of DNA double strand break repair by altNHEJ.

The V(D)J recombinase plays a role in directing which mode of repair occurs after cleavage. After cleavage, RAG retains association with the coding and signal ends in a post-cleavage complex, which is important for proper cNHEJ. Mice harboring a RAG-1 hypomorphic mutation S723C, which can catalyze cleavage but is defective in post-cleavage complex formation, exhibit aberrant DSB formation and predisposition toward chromosomal translocations in a p53-deficient background(107). Additionally non-consensus heptamer motifs, including a subset of

cryptic RSSs, can destabilize the post-cleavage complex and allow the DNA ends to be repaired by altNHEJ(203,205). This suggests a role for the RSS sequence in maintaining genomic stability. The C terminus of RAG-2 is important for proper DSB repair. Deletion the C-terminus allows altNHEJ to repair DSB in both cNHEJ proficient and cNHEJ deficient mice(206). altNHEJ is characterized by extensive end resection, junctional microhomolgies and translocations(85). Further support for the role the RAG-2 C-terminus in DSB repair pathways choice came from studies utilizing RAG-2 C terminal truncation mutants, which exhibited robust alternative end joining(207).

To limit DSBs at multiple loci within development, developing lymphocytes utilize a DNA damage response that suppresses RAG-1 and RAG-2 expression upon detection of a DSB(208). Furthermore, DSBs can initiate an ATM dependent DNA damage repair pathway which causes RAG-2 to relocalize outside of the nucleus to centrosomes(209). The RAG-2 core is sufficient for this export, and mutation of T490 prevents this nuclear export(209).

Class switch recombination represents another source of programmed DNA DSB at the IGH locus. This process translocates productive VDJ exons from their initial positions upstream of Cμ and C□ to positions upstream of the □, □ or □exons. Recombination involves formation of DNA DSB within a pair of IGH switch regions. These breaks are repaired by the cNHEJ machinery. Deficiencies in DNA double strand break repair factors are associated with translocations of c-myc to switch region junctions and increased usage of altNHEJ as evident from junctional microhomology(210). Thus, DSB associated with CSR may be joined with to an aberrant RAG-induced DSB to form a translocation(210).

Recurrent translocations in lymphoid malignancies

Balanced chromosomal translocations involving the Ig loci and oncogenes are frequently observed in B cell lymphomas. B cell development involves multiple processes that subject Ig genes to DNA cleavage such as V(D)J recombination, somatic hypermutation, and class switch recombination. DSB generated within Ig loci by these processes are then susceptible to fusion with other DSB produced within oncogenes. In humans, recurrent RAG-mediated translocations have been proposed to include the chr8:14 translocation involving IgH and c-myc in Burkitt lymphoma, the chr11;14 translocation involving IgH and bcl-1 found in mantle cell lymphomas, the chr14:18 translocation involving IgH and bcl-2 in follicular lymphoma and the chr1:14 translocation involving IgH and bcl-10 found in mucosa-associated lymphoid tissue(MALT) lymphoma(183). A frequently observed translocation involves the switch regions of IgH and the c-myc gene(211). These Ig:Myc translocations can arise from activation-induced cytidine deaminase (AID) dependent and independent mechanisms(211). A similar translocation between IgH and sequences downstream of c-myc is observed in B cell lymphomas from ATM deficient mice(212). This work suggested a role for ATM in preventing the formation of chromosome 12 dicentrics via recombination between IgH DSB and breaks downstream of c-myc(212). The frequent occurrence of B lymphoid tumors harboring c-myc translocations has suggested preferential rearrangement with c-myc. By replacing the c-myc coding sequence with n-myc, Alt and colleagues demonstrated that there are features of the c-myc locus that make it a preferred translocation target(213). Within the IgH locus, the regulatory region (IgH3'RR) is important for RAG initiated IgH-c-myc translocations, but not for class switch recombination initiated IgH-c-myc translocations(214).

Sequencing methodologies

With the rise of the genomics era, sequencing methodologies have assumed a central role in the analysis of biological phenomena. Detecting and quantifying genomic double stranded breaks (DSBs) has been a challenge, but much progress has been made because of advancements in sequencing technology. Analyses of short read paired-end sequences present many challenges in the analysis of cancer genomes. Roth and colleagues developed a method to identify tumor-specific structural variants that was designed to reduce false positives(215). They applied this method to analyzing a core RAG-2, p53 deficient mouse and were able to identify specific rearrangements with support of as few as 2 independent read pairs(215).

Alt and coworkers have developed a high-throughput, genome-wide translocation sequencing method, HTGTS, capable of identifying genome-wide translocation junctions arising from a specific cellular DSB created by the I-Sce1 meganuclease.(185). This method permits identification of other genomic DSB that are capable of joining to the test DSB. To test this method, DSB were created in the c-myc and IgH loci in activated B cells. The majority of translocations were intrachromosomal and frequently fell near transcription start sites (TSS)(185,216). The method was then expanded to other sources of programmed DSBs, such as Cas9 or TALENs, which provided useful data regarding off-target cutting by these engineered nucleases(217). Using RAG-generated breaks as bait, this technology was able to identify a large number of RAG “off-target” cleavage events(177).

A variety of methods, such as BLESS(218), have been developed to detect DSB, but most have only been efficacious in cell lines. Recently, Sleckman and colleagues developed sequencing methodologies, initially HCoDES and later END-Seq, that are capable of identifying DSB within the genome both in cell culture and in vivo(219,220). These methods utilize hairpin adapters followed by a pulldown to identify precise locations within the genome containing DSBs. This group examined RAG generated DSB and was able to detect over 200 off-target RAG mediated DSBs(219).

Approach

To characterize genomic rearrangements associated with mistimed V(D)J recombination *in vivo*, we employed a mouse model carrying the RAG-2 T490A mutation, which ablates the cyclinA/cdk2 phosphorylation site that triggers RAG-2 degradation at the G1-to-S transition. Expression of the RAG-2(T490A) mutant protein persists throughout the cell cycle. To reveal potential oncogenic effects, we the RAG-2^{T490A} allele onto a p53-deficient background(62). Of particular note, the T490A mutation is dominant.

We employ the RAG-2(T490A) p53-deficient mouse model of mistimed V(D)J recombination to (1) identify aberrant V(D)J recombination events in vivo, and (2) identify translocation partners involved in chromosomal rearrangements involving an antigen receptor locus. We developed a targeted method, based on the approach of Halper-Stromberg et al. (221), to selectively sequence V(D)J recombination events in this model (221). Several technologies exist to detect translocations within the genome. Of relevance, high throughput genomic translocation sequencing (HTGTS) is capable of detecting translocation partners by following a specific

targeted ‘bait’ double stranded break (185). Our technology has the advantage of being able to capture spontaneous RAG generated DSB within the genome and the following translocation partner in an unbiased fashion. We demonstrate that this method can detect canonical V(D)J rearrangements and translocations involving antigen receptor loci.

Results

Characterization of lymphoid tumors from mice bearing the dominant RAG-2(T490A) mutation on a p53-deficient background.

Previous work from this laboratory demonstrated that RAG-2(T490A) is oncogenic in the B and T lymphoid lineages when expressed on a p53-null background (62). With respect to the malignant phenotype RAG-2(T490A) is dominant, with no apparent differences between mice that are heterozygous or homozygous with respect to the RAG-2(T490A) allele (62). Spectral karyotypic analysis and fluorescence in situ hybridization revealed that lymphoid malignancies arising in RAG-2(T490A), p53-null mice exhibit complex chromosomal translocations involving antigen receptor loci (62). Here we employ hybridization-based capture sequencing to characterize genomic structural variants in these lymphoid malignancies at single-nucleotide resolution.

We began by collecting lymphoid tumors from p53-deficient mice homozygous or heterozygous for the RAG-2(T490A) allele (*RAG-2*$^{T490A/+}$*Trp53*$^{-/-}$ or *RAG-2*$^{T490A/T490A}$*Trp53*$^{-/-}$) and from p53-deficient control mice homozygous for the floxed, wild-type RAG-2 allele (*RAG-2*$^{fl/fl}$*Trp53*$^{-/-}$). . Among all three genotypes, from 70% percent of mice developed tumors by 24 weeks of age, with an average age of 23 weeks upon euthanization. Tumors were dispersed and the distributions of cell size and granularity were determined by flow cytometry; lymphoma cells exhibited increased size and granularity with respect to normal lymphocytes. Lymphomas were assayed for expression of the B and T lineage markers B220 and CD3, respectively (Fig. 1A). CD3^{+}, T lineage tumors were further assessed for surface expression of CD4 and CD8 (Fig. 1A).

Of the 45 *RAG-2*$^{T490A/+}$ *Trp53*$^{-/-}$ and *RAG-2*$^{T490A/T490A}$*Trp53*$^{-/-}$ mice analyzed, more than 82 percent developed splenic or thymic tumors (Fig. 1B). In about 18 percent of the *RAG-2*$^{T490A/+}$ *Trp53*$^{-/-}$ and *RAG-2*$^{T490A/T490A}$*Trp53*$^{-/-}$ mice tumors arose in sites other than thymus or spleen, in contrast to *RAG-2*$^{fl/fl}$*Trp53*$^{-/-}$ mice, in which more than 36 percent of tumors involved non-lymphoid organs, defined as not arising from the spleen or thymus and not expressing T or B lineage markers. Excluding mice without lineage information, about 74 and 72 percent of tumors arising in RAG-2(T490A), p53-null and wild-type RAG-2, p53 null mice were of the T lineage respectively (Fig1C). B lymphoid tumors were exclusively found in RAG-2(T490A), p53-null animals (Fig1C), consistent with previous results (62).

Capture sequencing strategy and probe design

By spectral karyotyping and fluorescence in situ hybridization it was previously shown that uncoupling of V(D)J recombination from the cell cycle is associated with chromosomal translocations involving antigen receptor loci. The resolution of these earlier studies, however, was too low to provide information regarding the mechanisms by which these translocations were formed. To define the corresponding recombinant junctions at the nucleotide sequence level we employed a capture sequencing strategy. To identify structural rearrangements that involved antigen receptor loci, we designed RNA capture probes across a selection of the antigen receptor loci. Of the seven murine antigen receptor loci , IGH, IGL, IGK, TRB, TRAD and TRG, we included IGH, TRB, and TRAD in the analyses because (1) translocations involving IGH and TRAD were previously detected in lymphoid tumors from RAG-2(T490A), p53-null mice and (2) the TRB locus is developmentally analogous to IGH. The IGH and TRB capture regions spanned the genomic intervals 114666978-114720181 on chromosome 12 and 41483000-

4158570 on chromosome 6, respectively, which included the D and J clusters of each locus. The TRAD capture region included the interval 54732943-54844073 on chromosome 14, spanning the Dδ and Jδ clusters, Cδ, the Jα cluster and Cα. The larger intervals spanning the V segments at each locus were excluded, as their presence would greatly increase the size of the capture region, thereby limiting the depth of sequencing. Consequently, we were unable to detect non-canonical V segment rearrangements except for those involving recombination partners within the capture intervals.

Two probe sets were constructed, a combined IGH/TRB probe set and a TRAD probe set. Probes were arrayed at five-fold tiling so that each nucleotide within the capture regions was capable of hybridizing to five different probes (Fig. 2A). Because the antigen receptor loci contain repetitive elements in the form of gene segments and recombination signal sequences, we did not perform repeat-masking when designing these probes.

Sequence analysis pipeline

To identify genomic rearrangements involving the capture regions we performed hybrid capture and paired end Illumina sequencing; reads were analyzed according to the pipeline outlined in Fig. 2B. Using the TRAD probe set for enrichment, we analyzed 16 tumors from *RAG-2*$^{T490A/+}$ *Trp53*$^{-/-}$ or *RAG-2*$^{T490A/T490A}$*Trp53*$^{-/-}$ mice, 1 tumor from a *RAG-2*$^{fl/+}$*Trp53*$^{-/-}$ and 1 tumor from a *RAG-2*$^{+/+}$ *Trp53*$^{-/-}$ mouse (Table 1). Using the IGH/TRB probe set, we analyzed 6 tumors from *RAG-2*$^{T490A/+}$ *Trp53*$^{-/-}$ or *RAG-2*$^{T490A/T490A}$*Trp53*$^{-/-}$ mice (Table 1). An average of 13.5M reads were obtained from each tumor sample using the TRAD probe set. Reads were aligned to the reference mouse genome, mm9, using the short read aligner BWA-MEM(222). On average, 90.3% of reads mapped to the capture region, indicating efficient enrichment by hybrid

capture. We then took the set of all paired reads and used samtools (223) to remove those in which the paired reads mapped in a proper pair, thereby enriching for paired reads spanning recombinant junctions. After removal of properly paired reads we obtained about 1.75M reads from each tumor, or about 12% of each original dataset; these were mapped to the reference genome. After mapping, split and discordant reads were extracted using samblaster (224) and examined using the Integrative Genomics Viewer (IGV) (225-227). Once in IGV we scanned for the occurrence of sequence pileups spanning putative recombinant junctions. We then proceeded to verify recombination events by PCR and Sanger sequencing, as described below; as few as two split reads were sufficient to identify verifiable rearrangements.

Canonical V(D)J recombination events

We expected to find products of canonical V(D)J rearrangement among the recombinant junctions mapped by the procedure described above. Examination of split and discordant reads identified 18 pileups consistent with canonical V(D)J recombination events. We characterized one such V(D)J rearrangement, VH3.8-DSP2-JH1, obtained from a B lymphoid lymphoma arising in a *RAG-2*$^{T490A/T490A}$*Trp53*$^{-/-}$ mouse using the IgH/TRB probe set. The genomic rearrangement was represented by one pair of discordant reads and three split reads (Fig. 3A). The presence of this rearrangement was verified by PCR amplification with primers specific for the V_H3.8 and J_H1 gene segments, followed by Sanger sequencing (Fig. 3B). The split read sequence was evidently the product of a V_H3.8-to-D_{SP2}-to-J_H1 rearrangement; the sequence interposed between the V_H and J_H segments is consistent with use of $D_{SP2.3}$, $D_{SP2.4}$ or $D_{SP2.6}$ (Fig. 3C). Non-templated insertions were detected at the V_H-to-D and D-to-J_H junctions, as is typical of V(D)J recombination (Fig. 3C).

Interchromosomal rearrangements

Having established that the capture sequencing protocol was able to detect canonical V(D)J recombination events, we proceeded to look for translocations involving the capture regions. The sequence analysis pipeline described above identified 575 putative genomic rearrangements among all tumors analyzed; of these 481 were consistent with interchromosomal translocations. One of these, a t(14:16) translocation involving the TCRA locus and the Roundabout Guidance Receptor 2 (Robo2) gene, was obtained from a T lymphoid tumor arising in a *RAG-2*$^{T490A/T490A}$*Trp53*$^{-/-}$ mouse using the TRAD probe set (Fig. 4A). Robo2 encodes a transmembrane receptor for the repulsive guidance protein Slit2 and is implicated in axon guidance and cell migration(228). This translocation, which was represented by 3 discordant read pairs and 6 split reads (Fig. 4A), was verified by PCR (Fig. 4B) and by Sanger sequencing (Fig. 4C). The chromosomal breakpoints at the translocation junction reside in the 5' flank of the TRAJ49 gene segment and the intron between exons 3 and 4 of Robo2 (Fig. 4C). The TRAJ49 breakpoint lies within the nonamer of the 5' flanking RSS, consistent with RAG-mediated cleavage at the nonamer-distal end of the heptamer, followed by partial resection of the resulting signal end. of TRAJ49. While deletion of nucleotides from signal ends is rare in physiologic V(D)J recombination, the resection of the signal end seen in this instance is in agreement with our previous finding that the RAG-2(T490A) mutation is associated with an increase in the frequency of nucleotide loss at RSS signal junctions(62). Within the Robo2 gene, the breakpoint is unaccompanied by a nearby cryptic RSS. No microhomology was observed at the junction, consistent with repair by classical NHEJ. We did not detect this translocation in any of the other tumors analyzed.

A t(12;15) translocation was obtained from a B lymphoid tumor arising in a *RAG-2*$^{T490A/T490A}$*Trp53*$^{-/-}$ mouse using the IGH/TRB probe set (Fig. 4D). This rearrangement is documented by 5 split reads. One breakpoint lies between the Cδ and Cγ3 exons in the IGH locus on chromosome 12; resides 744kb downstream Myc and 358 kb downstream of Pvt1 on chromosome 15. This rearrangement places Myc and Pvt1 downstream of the IgH D constant region, in a transcriptionally convergent orientation. The translocation was verified by PCR and Sanger sequencing (Fig. 4E and F). The breakpoint on chromosome 15 is apparently unaccompanied by a cryptic RSS, while the breakpoint on chromosome 12 is flanked by a nearly canonical RSS (Fig 4F). The structure of the chromosome 12 breakpoint is consistent with RAG-mediated cleavage at the cryptic heptamer, followed by resection to the right of the cleavage site as depicted in Fig. 4F. It is possible that RAG cleaved downstream of the GTG of the heptamer, there was some nucleotide loss, and then joining of this segment to an available DNA end in Ch15. Minimal (1 bp) microhomology was observed at the translocation junction, raising the possibility of repair by microhomology-mediated end joining.

Overall landscape of structural variation

To obtain a more comprehensive view of structural variation in the tumors analyzed here we employed LUMPY, a tool for the probabilistic discovery of genomic structural variants in multiple samples (229). LUMPY analysis was performed on sequences obtained from 21 *RAG-2*$^{T490A/+}$ *Trp53*$^{-/-}$ or *RAG-2*$^{T490A/T490A}$*Trp53*$^{-/-}$ tumors, 1 *RAG-2*$^{fl/+}$*Trp53*$^{-/-}$, and 1 *RAG-2*$^{+/+}$*Trp53*$^{-/-}$, tumors. Of the sequences from *RAG-2*$^{T490A/+}$ *Trp53*$^{-/-}$ or *RAG-2*$^{T490A/T490A}$*Trp53*$^{-/-}$ tumors, 15 were obtained with the TRAD probe set and 6 with the IGH/TRB probeset (Fig 5A). Sequences from the *RAG-2*$^{fl/+}$*Trp53*$^{-/-}$ and *RAG-2*$^{+/+}$*Trp53*$^{-/-}$ tumors were obtained the TRAD probeset.

Duplications, inversions, deletions, and interchromosomal rearrangements were scored. An overall average of 54 structural rearrangements were detected per tumor sample. The majority of structural rearrangements detected by LUMPY, 481 of 575, or 84 percent, were interchromosomal rearrangements, consistent with translocation being the dominant form of structural variation within these tumors. Other than the presence of B lineage tumors arising solely in the RAG-2^{T490A} *Trp53*$^{-/-}$ tumors, striking differences were not observed between RAG-2^{T490A} *Trp53*$^{-/-}$ tumors and control mice.

Discussion

V(D)J recombination is a major source of oncogenic genomic variation in lymphoid cells. Indeed, aberrant V(D)J recombination has been estimated to underlie translocations in 40 - 70% of human pro- and pre-B cell malignancies (192,230). Such translocations may involve either (1) legitimate RAG cleavage at an Ig/TCR gene segment and illegitimate RAG cleavage at a proto-oncogene, or (2) legitimate RAG cleavage at a pair of Ig/TCR gene segments and cleavage at a proto-oncogene by a RAG-independent mechanism (192). We have employed an in vivo model of mistimed V(D)J recombination in the mouse as a source of RAG-associated translocations. The coupling of V(D)J recombination to the cell cycle is dependent on phosphorylation of RAG-2 at T490 and its subsequent ubiquitylation and degradation (59-62). This coupling directs RAG-generated DNA DSBs toward repair by NHEJ. Previous work from this laboratory has demonstrated that mice containing the RAG-2 T490A mutation on a p53-null background develop lymphoid malignancies characterized by complex, clonal chromosomal translocations, often involving antigen receptor loci(62).

In this chapter, we characterized genomic rearrangements obtained from lymphoid tumors arising in this mouse model. Our approach for interrogating V(D)J rearrangements and RAG-mediated structural variation involved targeted, antigen receptor gene-specific hybrid capture combined with Illumina sequencing(221). We confirm previous findings of this laboratory by showing that in T lineage and B lineage lymphomas from RAG-2(T490A), p53-null mice chromosomal translocations are the predominant form of genomic variation. We extend these findings by identifying these rearrangements at genome-wide scale and defining them at the nucleotide level(62).

The novelty of our approach lies in its ability to detect endogenously arising translocations in a mouse model of mistimed V(D)J recombination. High-throughput, genome wide translocation sequencing (HTGTS) has been used characterize translocations arising after introduction of non-physiologic DNA DSBs as bait (185,216). Another approach has used hairpin adapters to identify DNA double stranded breaks, regardless of the source of breakage, within the genome both in cell culture and in vivo (219,220). This methodology has the advantage of being able to detect DNA double stranded breaks in vivo. Our in vivo model and sequencing methodology are able to capture and enrich for endogenous, RAG-mediated structural variants. A similar panel called myTYPE has been applied clinically to analyze samples from patients with multiple myeloma; myTYPE can identify translocations involving the IgH locus(231). Our general approach may therefore have utility in the clinic.

Results presented above suggest that chromosomal translocation is the predominant type of structural variation in tumors from RAG-2(T490A), p53-null mice. This is in contrast to an

analysis of core RAG-2, p53-null mice in which most reported structural variants were deletions (199,200). We posit that this difference can be attributed to the sensitivity of our system in comparison to whole genome sequencing studies, the restriction of our analysis to rearrangements involving the capture regions, functional differences between the distinct RAG-2 mutants employed, or some combination of the above. With regard to the first possibility, our capture method may enrich for interchromosomal rearrangements whereas these potentially infrequent rearrangements are missed within a whole genome sequencing dataset because of the rarity of the event or the heterogeneity of the tumor sample. If this possibility is true, it is possible that intrachromosomal rearrangements are not the predominant form of structural variation in these tumors, but just the predominant source identified within the limitations of our methodology. In addition, the filtering method that we used, with which we excluded properly paired reads, may have eliminated many small deletions that could have been scored as proper pairs.

It remains possible that the differences between our model and the core RAG-2 model with respect to structural variation reflect differences between the RAG-2 phenotypes. Core RAG-2 lacks regulatory the inhibitory domain, the PHD, the degradation domain and the nuclear localization region (7,8,17,59-62,199,200). The PHD mediates allosteric activation of RAG by H3K4me3 (17,20,117). Disruption of the inhibitory domain renders RAG activity independent of H3K4me3 binding(54). By eliminating these domains localization or activity of RAG could be misregulated. It is therefore difficult to know whether the structural rearrangements observed result from alternative DNA double strand break repair machinery or from another misregulated

RAG activity. Our model allows us to examine the effects of RAG induced double stranded breaks outside of G1 and to assess these effects on genomic integrity.

We did not observe many recurrent genomic rearrangements between tumor samples. This was in contrast to core RAG-2, Trp53$^{-/-}$ mice which exhibited recurrent deletions in Notcht1, Pten, Ikzf1, Jak1, Phlda1, Trat1, and Agpat9 genes (199). This difference may result from undercounting of deletions by our method or by differences in regulation of the RAG-2 variants used. Previous work from this laboratory employed spectral karyotyping and fluorescence in situ hybridization to identify translocations TRAD in T lineage tumors from RAG-2(T490A), p53-null mice(62). The capture sequence method described here also identified a translocation involving TRAD. Analysis of the translocation junction at the nucleotide level implicates DNA cleavage by RAG at TRAJ49 within the J cluster and a RAG-independent cleavage event within the Robo2 locus. Moreover, the apparent resection of the TRAJ49 signal end is characteristic of aberrant signal joints in mice expressing RAG-2(T490A) (62). Our results are consistent those of Halper-Stromberg and colleagues, who, employing a similar capture-based approach, failed to observe recurrent translocations in cancer cell lines or human primary lymphoid tumor samples (221).

As stated above, the ch14:16 translocation between TRA and Robo2 that we identified in T lineage thymic tumor is likely a type 2 event. Although a minimal CAC motif is present about 20 nt from the breakpoint, extensive end resection in the incorrect direction would have had to occur prior to ligation to form the translocation junction that we observed. We favor the interpretation that this rearrangement arose from the joining of a RAG-mediated signal end at TRAJ49 to a RAG-independent breakpoint at the Robo2 locus. In accordance with the ability of

RAG-2(T490A) to support accumulation of signal ends outside of G0/G1, a signal end could have illegitimately been joined to Robo2 giving rise to the structural variant we observed (62). The location of Robo2 in nuclear space has not been examined, so it is not known whether these two loci colocalize.

The ch12:15 translocation between IgH and downstream of Myc identified in the B lineage thymic tumor 2008 is also likely to have arisen from a type 2 translocation, as there is limited evidence for a canonical or cryptic RSS surrounding the ch15 breakpoint suggesting the Ch15 breakpoint may not have been RAG-mediated. Additionally, the breakpoint within the IgH locus falls between the Cδ and Cγ3 constant regions but does not overlap with a known switch region. Translocations involving Myc and the IgH locus have been documented in healthy patient thymus cells (232,233). Additionally, it has been shown that Myc and IgH occupy the same location within the nuclear space (174).

Overall, we present a new method to identify structural variation involving antigen receptor loci. We employ this approach to characterize structural variants associated with mistimed RAG activity. In so doing we identify endogenous V(D)J recombination events as well as translocations involving antigen receptor loci. This methodology can be applied to a variety of tumor samples and the analysis utilizes free bioinformatic tools, making this methodology easily adoptable by others. The presence of chromosomal translocations involving TRAD are reminiscent of those observed in 20% of human T cell acute lymphoblastic leukemia, suggesting that these results may shed light on mechanisms underlying this disorder (192).

Summary and Future Directions

V(D)J recombination is coupled to the cell cycle through the periodic destruction of RAG-2 by phosphorylation at T490 and subsequent degradation via the ubiquitin-proteasome pathway. The T490A mutation eliminates this phosphorylation site and allows for constitutive RAG expression and RAG-generated DNA double stranded breaks throughout the cell cycle. To assess the consequences of this uncoupling, we utilized a mouse model containing the T490A mutation crossed into a p53 deficient background. These mice exhibited complex chromosomal translocations involving antigen receptor loci. To better define the structural variation and genomic instability within this model, we employed a capture sequencing-based methodology capable of detecting endogenous V(D)J rearrangements and novel translocations involving antigen receptor loci.

Since p53 deficient tumors have been well characterized and demonstrated not to exhibit translocations, we only sequenced two control tumors using our capture-sequencing methodology. However, since we do not have a large sample size of control tumors, it remains difficult to draw clear conclusions regarding mistimed RAG activity and structural variation. To expand upon this work, we could sequence additional control tumors to draw clearer conclusions regarding which rearrangements are specific to the mistimed V(D)J recombination model. Another limitation of this study was the small sample size of tumors analyzed using the IGH/TRB probe set. Analyzing more tumors with this probe set would be informative and may provide us with additional insight into recurrent translocations and rearrangements.

Since spatial proximity of loci may play a role in translocations, it would also be interesting to culture the tumor cells and perform Hi-C analyses to see if loci involved in translocations are located close to each other in nuclear space. Previous studies have identified colocalization of the IgH locus and c-myc.

Since all of our analyses were performed on bulk tumor cells, it is possible that we are missing some of the rarer structural variants. Future studies could use FACS to isolate the tumor cells from the non-tumor cells, and then perform the capture sequencing methodology. This would provide an internal control within an animal and could identify differences in structural variation between tumor and non-tumor tissue. Additionally, once technologies advance, single cell analyses utilizing this probe set and pipeline would be very interesting to identify potential recurrent translocations, but this is currently technically challenging due to the large amount of DNA needed to perform the pulldown.

Occasionally the RAG-2(T490A), p53-deficient mice developed non-lymphoid tumors. Our analyses focused primarily on lymphoid tumors, but future studies could perform histology on these non-lymphoid tumors to identify the underlying pathology.

Figures and Legends

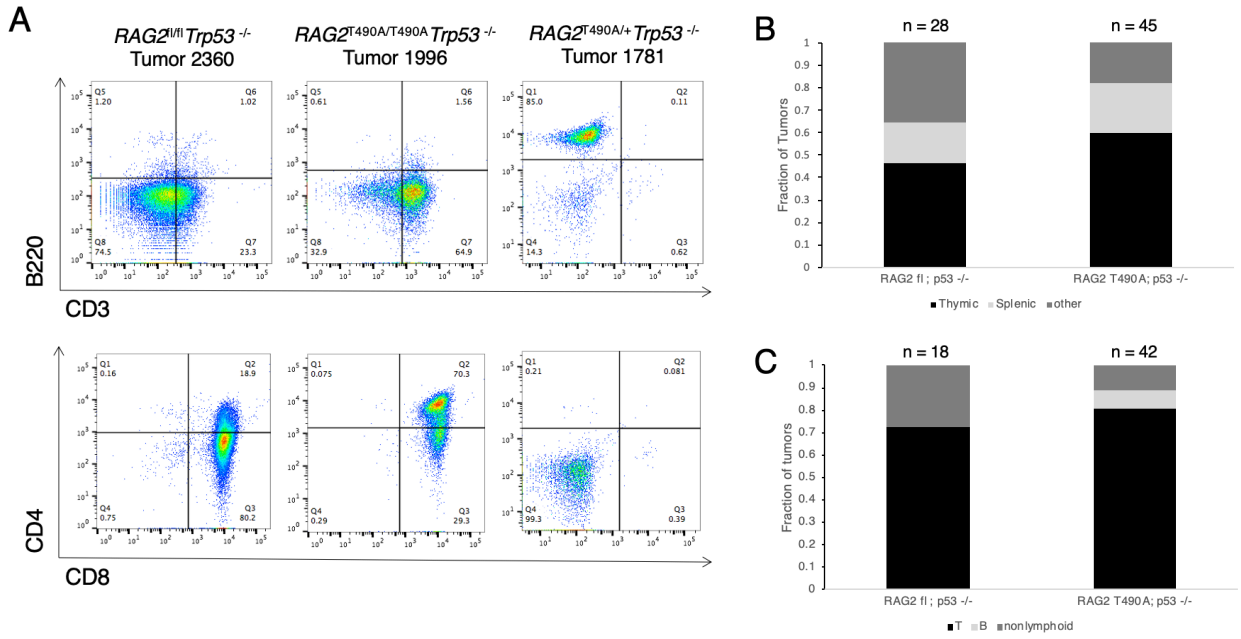


Figure 1: RAG-2 (T490A); p53 -/- mice develop T and B lineage lymphoid tumors. A, *above* representative flow cytometry plots of thymic tumors obtained from RAG-2(fl); $p53^{-/-}$ controls or RAG-2 (T490A); $p53^{-/-}$, genotype and mouse number indicated above, stained for B and T lineage markers, B220 and CD3 respectively. Gates set based on unstained and stained WT C57BL/6J mouse thymus controls. *Below* flow cytometry plots stained for T cell markers CD4 and CD8. Gates set based on same controls as above. B, Types of malignancies generated by genotype with sample size indicated above. Sample size is indicated above. Other category includes any malignancy other than spleen or thymus, i.e. enlarged liver, other solid tumors, etc… C, Lymphoid lineage of thymic and splenic tumors plotted by genotype. Phenotypes were determined by flow cytometry using the markers indicated in Fig1A. Sample size is indicated above.

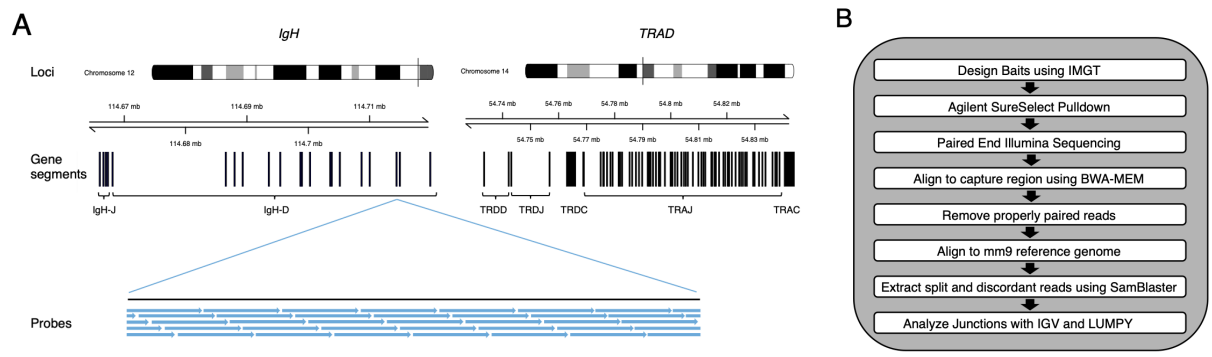


Figure 2: Design of custom antigen receptor capture. A, Schematic of the biotinylated RNA probe design. Probes were designed to span the J and D segments of IgH locus, the D and J segments of the TRA and TRD loci. *Above* Ideogram of mouse chromosomes 12 and 14, with antigen receptor locus location indicated by a vertical line. *Middle* Annotated Immunogenetics Information System (IMGT) antigen receptor genes plotted by genomic location. *Below* Enlarged view to demonstrate 5x tiling of RNA probes, illustrated as blue arrows, across the capture region. B, Outline of capture sequencing and analysis pipeline beginning with probe design, through structural analyses.

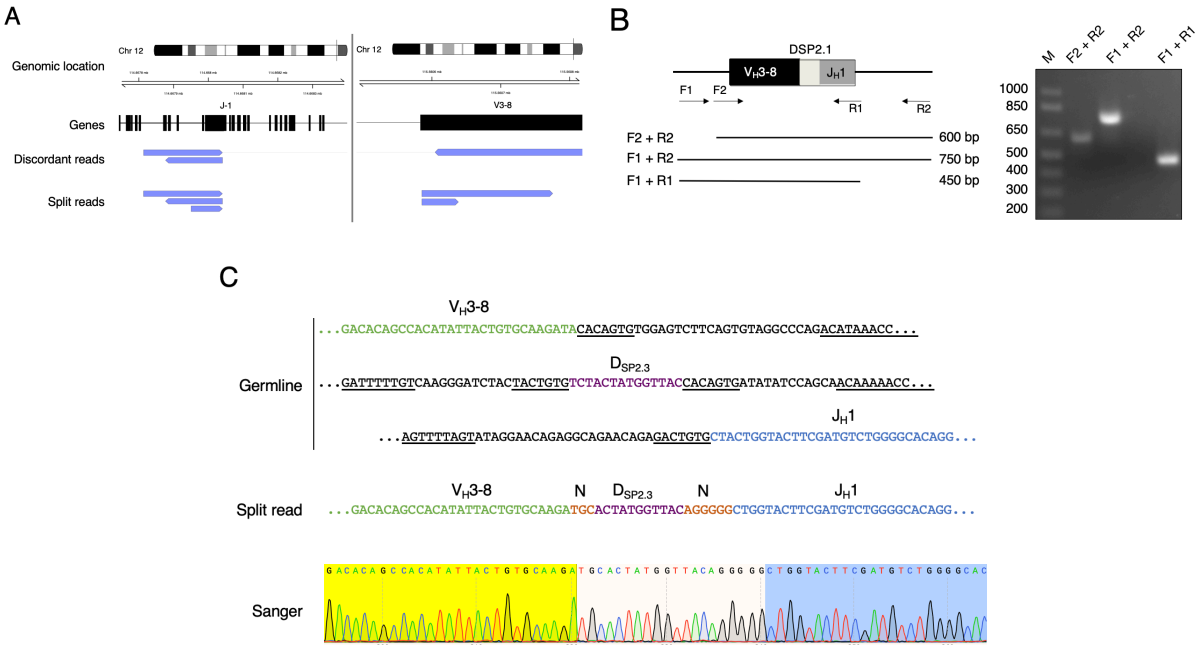


Figure 3: Antigen receptor capture sequencing is able to efficiently identify canonical V(D)J rearrangements. A, Example of a V(D)J rearrangement identified in thymic tumor 2008. *Above* Ideogram of mouse chromosome 12, with IgH locus location indicated by a vertical line. *Middle* IMGT known gene IgH J-1 genomic location is plotted and *Below* discordant and split reads, indicated in purple, are aligned to the mm9 mouse reference genome. A discordant read pair is connected by a light grey line. *Left* highlights IgH J-1 location, and *right* highlights IgH V3-8 location. B, PCR verification of thymic tumor 2008 V(D)J rearrangement. *Left*, schematic of a successfully rearranged locus with primers indicated as black arrows. Primer combinations and predicted product sizes are illustrated below. *Right*, agarose gel of PCR products with primer pairs indicated above the lanes. C, Schematic of an example split read covering the V(D)J junction. *Above track,* a selection of genomic sequence color coded with blue and green being included in the split read sequence, black being the flanking genomic sequence, and orange

indicating recombination signal sequence (RSS). *Middle tracks*, linear schematic of genomic unrearranged IgH locus followed underneath by a zoomed in version of the rearranged locus identified within the split read, color coded with V3-8 in blue, DSP2.3 in purple, J-1 in green, and non-templated nucleotides in red. Below portion of the split read sequence spanning V3-8, DSP2.3, and J-1 with same color coding as schematic. *Below track,* portion of sanger sequencing read confirming V(D)J rearrangement. V3-8 is highlighted in blue.

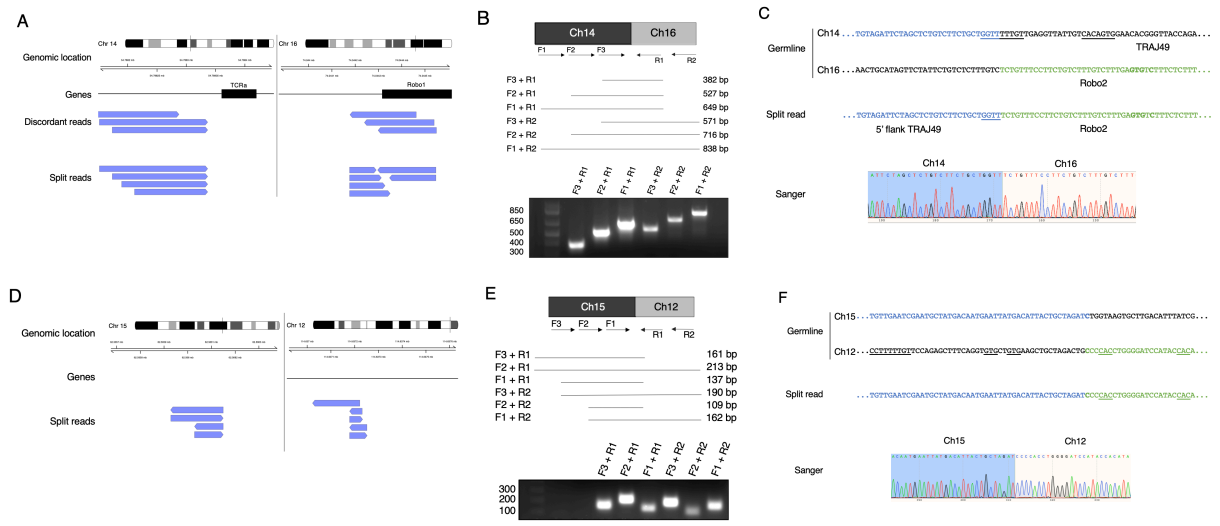


Figure 4: Antigen receptor capture sequencing permits characterization of translocations in RAG-2(T490A);p53-/- thymic tumors involving antigen receptor loci. A, Example of a translocation identified in thymic tumor 1931 involving an antigen receptor locus. *Above* Ideogram of mouse chromosome 14 and 16, with genomic locations involved in rearrangement indicated by vertical lines. *Middle* UCSC known gene genomic locations are plotted and *Below* discordant and split reads, indicated in purple, are aligned to the mm9 mouse reference genome. *Left* highlights Ch14 involvement in rearrangement, and *right* highlights Ch16 involvement in rearrangement. B, PCR verification of thymic tumor 1931 translocation. *Above*, schematic of a Ch14:16 translocation with primers indicated as black arrows. Primer combinations and predicted product sizes are illustrated below. *Below*, agarose gel of PCR products with primer pairs indicated above the lanes. C, Schematic of an example split read covering the Ch14:16 translocation. *Above track,* a selection of genomic sequence color coded with blue and green

being included in the split read sequence, black being the flanking genomic sequence, and bold underline indicating recombination signal sequence (RSS), and bold indicating potential RSS-like motifs. *Middle tracks*, linear schematic of genomic unrearranged Ch14 and 16 loci followed underneath by a zoomed in version of the rearrangement identified within the split read, color coded with Ch14 in blue and Ch16 in green. *Below* portion of sanger sequencing read confirming Ch14:16 translocation. Ch14 is highlighted in blue. D, Example of a Ch15:12 translocation identified in thymic tumor 2008. *Above* Ideogram of mouse chromosome 15 and 12, with genomic locations involved in rearrangement indicated by vertical lines. *Middle* UCSC known gene genomic locations are plotted, and *Below* discordant read, indicated in purple, are aligned to the mm9 mouse reference genome. *Left* highlights Ch15 involvement in rearrangement, and *right* highlights Ch12 involvement in rearrangement. E, PCR verification of thymic tumor 2008 Ch15:12 translocation. *Above*, schematic of a Ch15:12 translocation with primers indicated as black arrows. Primer combinations and predicted product sizes are illustrated below. *Below*, agarose gel of PCR products with primer pairs indicated above the lanes. F, Schematic of an example split read covering the Ch15:12 translocation. *Above track,* a selection of genomic sequence color coded with blue and green being included in the split read sequence, black being the flanking genomic sequence, and underline indicating recombination signal sequence (RSS)-like sequence and bold indicating potential microhomology. *Middle tracks*, linear schematic of genomic unrearranged Ch15 and 11 loci followed underneath by a zoomed in version of the rearrangement identified within the split read, color coded with Ch15 in blue and Ch12 in green. *Below* portion of sanger sequencing read confirming Ch15:12 translocation. Ch15 is highlighted in blue.

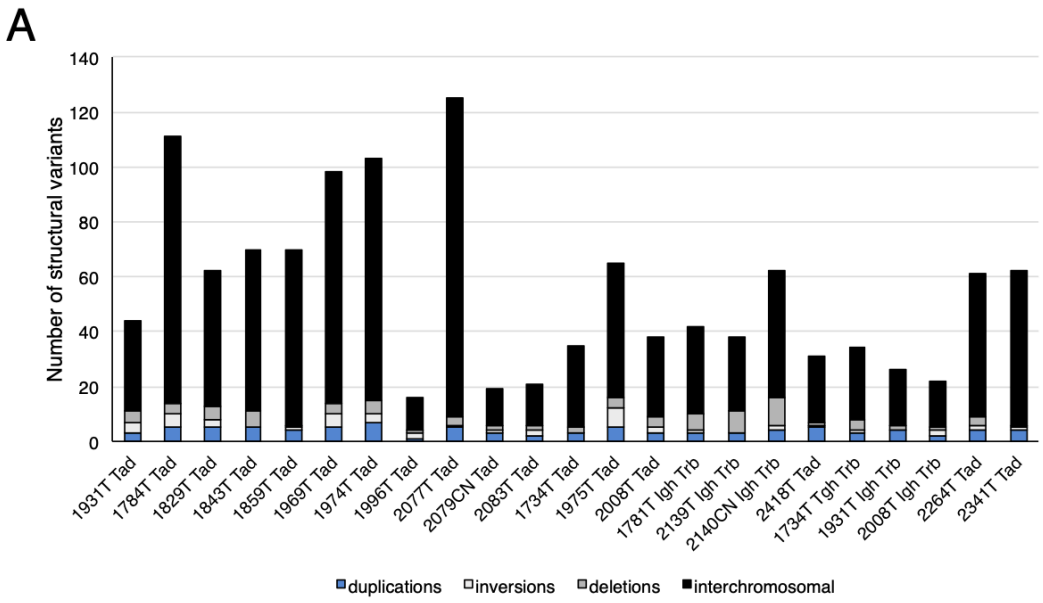


Figure 5: LUMPY structural variation analyses reveal interchromosomal rearrangements in RAG-2(T490A); p53-/- mouse tumors. A, graph of types of structural rearrangements identified by LUMPY analysis. RAG-2 (T490A);p53$^{-/-}$ tumors are indicated below by tumor number, type of tumor T thymic tumor or CN cervical node tumor, and probeset used for capture, either TRAD probe set labeled Tad or IgH/TRB probe set labeled Igh Trb. Asterisks indicate RAG-2(fl);p53$^{-/-}$ controls.

Mouse Tumor #	Genotype	Tumor location	Lineage	Probe set utilized
1784	RAG2 T490A/+; p53 -/-	Thymus	T	TRAD
1996	RAG2 T490A/T490A; p53 -/-	Thymus	T	TRAD
2077	RAG2 T490A/T490A; p53 -/-	Thymus	T	TRAD
2079	RAG2 T490A/T490A; p53 -/-	Cervical Node	T	TRAD
2083	RAG2 T490A/T490A; p53 -/-	Thymus	T	TRAD
1974	RAG2 T490A/+; p53 -/-	Thymus	T	TRAD
1734	RAG2 T490A/+; p53 -/-	Thymus	T	IgH/TRB, TRAD
1931	RAG2 T490A/T490A; p53 -/-	Thymus	T	IgH/TRB, TRAD
2008	RAG2 T490A/T490A; p53 -/-	Thymus	T	IgH/TRB, TRAD
2139	RAG2 T490A/T490A; p53 -/-	Thymus	T	IgH/TRB, TRAD
2140	RAG2 T490A/T490A; p53 -/-	Cervical Node	B	IgH/TRB
1781	RAG2 T490A/+; p53 -/-	Thymus	B	IgH/TRB
2264	RAG2 WTfl/+; p53 -/-	Thymus	T	TRAD
2341	RAG2 +/+; p53 -/-	Thymus	T	TRAD
1975	RAG2 T490A/T490A; p53 -/-	Thymus	T	TRAD
2418	RAG2 T490A/T490A; p53 -/-	Thymus	T	TRAD
1969	RAG2 T490A/T490A; p53 -/-	Thymus	T	TRAD
1843	RAG2 T490A/+; p53 -/-	Thymus	T	TRAD
1859	RAG2 T490A/T490A; p53 -/-	Thymus	T	TRAD
1829	RAG2 T490A/+; p53 -/-	Thymus	T	TRAD

Table 1: Mouse tumors analyzed.

Experimental Methods

Mouse generation and handling.

RAG-2(T490A) mice (62) and RAG-2 (WT*) controls containing a floxed neomycin resistance marker were crossed onto a p53-/- deficient background (Jackson Laboratories) to generate RAG-2(T490A) p53 -/- mice and RAG-2(WT*) p53 -/- mice. Mice were genotyped utilizing primers from Zhang 2011. All mice were maintained in a pathogen free environment. Mice were monitored daily for signs of distress, and then immediately euthanized, dissected, and tumor tissue was isolated. Tumor tissue was frozen in liquid nitrogen for further analysis, cultured in RPMI in preparation for metaphase analysis, and immunophenotyped by flow cytometry. Cells were stained with APC-CD3, V450-B220, PE-CD4, and FITC-CD8 antibodies (Becton Dickinson). All flow cytometry utilized a WT 6-8 week old mouse thymus or spleen for gating controls.

SureSelect library design.

Antigen receptor loci genomic locations in the mouse reference genome mm10 were downloaded from the international ImMunoGeneTics (IMGT) database and converted to mm9 using UCSC genome browser's LiftOver tool. We designed two sets of probes, IgH/TRB and TRA/D. For each antigen receptor locus, we chose to cover the D and J segments plus a 200 bp target window immediately flanking the D and J segments. For the IgH/TRB set, we tiled across IgHD1 through IgHJ4 and TRBD1 through TRBV31with 200bp extra on each end (mm9 coordinates: chr6:41483000-4158570, chr12:114666978-114720181). We utilized 3278 120-mer baits, with a total size of 78.792kbp. For the TRAD probeset we tiled across TRDD1 through TRAC, with 200bp extra on each end (mm9 coordinates: chr14:54732943-54844073). We utilized 4657 120-

mer baits, with a total size of 111.144kbp. We employed five-fold Agilent SureSelect bait tiling, with no masking. These baits were used to pull-down these antigen receptor sequences using eArray.

Targeted DNA library preparation and sequencing.

High quality genomic DNA was obtained from either fresh or frozen tumor tissue according to GenElute Mammalian Genomic DNA Miniprep Kit manufacturer instructions (Sigma G1N10). The Johns Hopkins Genetic Resources High Throughput Sequencing core facility performed library preparation, capture, and sequencing. Each gDNA sample was fragmented to 500bp fragments using a Covaris E210 system. Sample DNA was end-repaired using, purified via SPRI bead cleanup, dA-tailed, and ligated to index-specific, paired end adapters from Illumina. Agilent's SureSelect Target Enrichment System Kit for Illumina Paired-End Multiplexed Sequencing was utilized for library preparation and pull-down. 500ng of library was hybridized to the custom biotinylated RNA oligonucleotides. Magnetic bead selection was performed to capture the RNA-DNA hybrids and RNA was digested according to Agilent's protocol. Remaining captured DNA capture was PCR-amplified. The multiplexed libraries were then sequenced on Illumina HiSeq generating an average of 15-16M 250bp paired-end reads across the captured region.

Rearrangement identification.

We first aligned reads to the capture region using BWA-MEM with standard settings(222). On average, about 90% of reads mapped to the capture region. Next, utilizing Samtools we removed properly paired reads, defined as containing a proper orientation, standard insert size, and both

pairs mapping to the capture region(222,223,234). We took the remaining reads, which averaged between 2-2.5M reads, and aligned the reads to mm9. On average 99% of these reads mapped to mm9. Once mapped, we extracted split and discordant reads using Samblaster and analyzed these reads by visualization in Integrative Genomics Viewer(IGV) and by structural variation analyses using LUMPY(224,229).

Validation.

We designed primers to amplify across rearrangement breakpoints. We considered an event validated if we observed the expected DNA fragment within the sample but not in the control DNA. Additionally, we sanger sequenced these PCR products to confirm the exact sequence of the breakpoint junction. We validated both canonical V(D)J rearrangements and structural variations in the form of novel translocations.

References

1. Gellert, M. (2002) V(D)J recombination: RAG proteins, repair factors, and regulation. *Annual Review of Biochemistry* **71**, 101-132
2. Fugmann, S. D. (2010) The origins of the Rag genes--from transposition to V(D)J recombination. *Semin Immunol* **22**, 10-16
3. Schatz, D. G., and Swanson, P. C. (2011) V(D)J recombination: mechanisms of initiation. *Annual Review of Genetics* **45**, 167-202
4. Deriano, L. R., DB. (2013) Modernizing the Nonhomologous End-Joining Repertoire: Alternative and Classical NHEJ Share the Stage. *Annu Rev Genet* **47**, 433-455
5. Kim, M. S., Lapkouski, M., Yang, W., and Gellert, M. (2015) Crystal structure of the V(D)J recombinase RAG1-RAG2. *Nature* **518**, 507-511
6. Liu, Y., Subrahmanyam, R., Chakraborty, T., Sen, R., and Desiderio, S. (2007) A plant homeodomain in RAG-2 that binds Hypermethylated lysine 4 of histone H3 is necessary for efficient antigen-receptor-gene rearrangement. *Immunity* **27**, 561-571
7. Matthews, A. G., Kuo, A. J., Ramon-Maiques, S., Han, S., Champagne, K. S., Ivanov, D., Gallardo, M., Carney, D., Cheung, P., Ciccone, D. N., Walter, K. L., Utz, P. J., Shi, Y., Kutateladze, T. G., Yang, W., Gozani, O., and Oettinger, M. A. (2007) RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. *Nature* **450**, 1106-1110
8. Ramon-Maiques, S., Kuo, A. J., Carney, D., Matthews, A. G., Oettinger, M. A., Gozani, O., and Yang, W. (2007) The plant homeodomain finger of RAG2 recognizes histone H3 methylated at both lysine-4 and arginine-2. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 18993-18998
9. Van Ness, B. G., Weigert, M., Coleclough, C., Mather, E. L., Kelley, D. E., and Perry, R. P. (1981) Transcription of the unrearranged mouse C kappa locus: sequence of the initiation region and comparison of activity with a rearranged V kappa-C kappa gene. *Cell* **27**, 593-602
10. Yancopoulos, G. D., and Alt, F. W. (1985) Developmentally controlled and tissue-specific expression of unrearranged VH gene segments. *Cell* **40**, 271-281
11. Chakraborty, T., Chowdhury, D., Keyes, A., Jani, A., Subrahmanyam, R., Ivanova, I., and Sen, R. (2007) Repeat organization and epigenetic regulation of the DH-Cmu domain of the immunoglobulin heavy-chain gene locus. *Molecular cell* **27**, 842-850
12. Goldmit, M., Ji, Y., Skok, J., Roldan, E., Jung, S., Cedar, H., and Bergman, Y. (2005) Epigenetic ontogeny of the Igk locus during B cell development. *Nature immunology* **6**, 198-203
13. Jung, D., Giallourakis, C., Mostoslavsky, R., and Alt, F. W. (2006) Mechanism and control of V(D)J recombination at the immunoglobulin heavy chain locus. *Annual Review of Immunology* **24**, 541-570
14. Morshead, K. B., Ciccone, D. N., Taverna, S. D., Allis, C. D., and Oettinger, M. A. (2003) Antigen receptor loci poised for V(D)J rearrangement are broadly associated with BRG1

and flanked by peaks of histone H3 dimethylated at lysine 4. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 11577-11582
15. Subrahmanyam, R., Du, H., Ivanova, I., Chakraborty, T., Ji, Y., Zhang, Y., Alt, F. W., Schatz, D. G., and Sen, R. (2012) Localized epigenetic changes induced by DH recombination restricts recombinase to DJH junctions. *Nature immunology* **13**, 1205-1212
16. Callebaut, I., and Mornon, J. P. (1998) The V(D)J recombination activating protein RAG2 consists of a six-bladed propeller and a PHD fingerlike domain, as revealed by sequence analysis. *Cellular and molecular life sciences : CMLS* **54**, 880-891
17. Lu, C., Ward, A., Bettridge, J., Liu, Y., and Desiderio, S. (2015) An autoregulatory mechanism imposes allosteric control on the V(D)J recombinase by histone H3 methylation. *Cell reports* **10**, 29-38
18. Grundy, G. J., Ramon-Maiques, S., Dimitriadis, E. K., Kotova, S., Biertumpfel, C., Heymann, J. B., Steven, A. C., Gellert, M., and Yang, W. (2009) Initial stages of V(D)J recombination: the organization of RAG1/2 and RSS DNA in the postcleavage complex. *Mol Cell* **35**, 217-227
19. Shimazaki, N., Tsai, A. G., and Lieber, M. R. (2009) H3K4me3 stimulates the V(D)J RAG complex for both nicking and hairpinning in trans in addition to tethering in cis: implications for translocations. *Molecular cell* **34**, 535-544
20. Bettridge, J., Na, C. H., Pandey, A., and Desiderio, S. (2017) H3K4me3 induces allosteric conformational changes in the DNA-binding and catalytic regions of the V(D)J recombinase. *Proc Natl Acad Sci U S A* **114**, 1904-1909
21. Schatz, D. G., and Baltimore, D. (1988) Stable expression of immunoglobulin gene V(D)J recombinase activity by gene transfer into 3T3 fibroblasts. *Cell* **53**, 107-115
22. Schatz, D. G., Oettinger, M. A., and Baltimore, D. (1989) The V(D)J recombination activating gene, RAG-1. *Cell* **59**, 1035-1048
23. Oettinger, M. A., Schatz, D. G., Gorka, C., and Baltimore, D. (1990) RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. *Science* **248**, 1517-1523
24. Oettinger, M. A., Stanger, B., Schatz, D. G., Glaser, T., Call, K., Housman, D., and Baltimore, D. (1992) The recombination activating genes, RAG 1 and RAG 2, are on chromosome 11p in humans and chromosome 2p in mice. *Immunogenetics* **35**, 97-101
25. Shinkai, Y., Rathbun, G., Lam, K. P., Oltz, E. M., Stewart, V., Mendelsohn, M., Charron, J., Datta, M., Young, F., Stall, A. M., and et al. (1992) RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement. *Cell* **68**, 855-867
26. Mombaerts, P., Iacomini, J., Johnson, R. S., Herrup, K., Tonegawa, S., and Papaioannou, V. E. (1992) RAG-1-deficient mice have no mature B and T lymphocytes. *Cell* **68**, 869-877
27. Schatz, D. G. (1999) Transposition mediated by RAG1 and RAG2 and the evolution of the adaptive immune system. *Immunol Res* **19**, 169-182
28. Agrawal, A., Eastman, Q. M., and Schatz, D. G. (1998) Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature* **394**, 744-751
29. Tsai, C. L., and Schatz, D. G. (2003) Regulation of RAG1/RAG2-mediated transposition by GTP and the C-terminal region of RAG2. *Embo j* **22**, 1922-1930
30. Brandt, V. L., and Roth, D. B. (2009) Recent insights into the formation of RAG-induced chromosomal translocations. *Adv Exp Med Biol* **650**, 32-45

31. Jiang, H., Ross, A. E., and Desiderio, S. (2004) Cell cycle-dependent accumulation in vivo of transposition-competent complexes between recombination signal ends and full-length RAG proteins. *J Biol Chem* **279**, 8478-8486
32. Chen, X., Cui, Y., Wang, H., Zhou, Z. H., Gellert, M., and Yang, W. (2020) How mouse RAG recombinase avoids DNA transposition. *Nat Struct Mol Biol* **27**, 127-133
33. Carmona, L. M., Fugmann, S. D., and Schatz, D. G. (2016) Collaboration of RAG2 with RAG1-like proteins during the evolution of V(D)J recombination. *Genes Dev* **30**, 909-917
34. Huang, S., Tao, X., Yuan, S., Zhang, Y., Li, P., Beilinson, H. A., Zhang, Y., Yu, W., Pontarotti, P., Escriva, H., Le Petillon, Y., Liu, X., Chen, S., Schatz, D. G., and Xu, A. (2016) Discovery of an Active RAG Transposon Illuminates the Origins of V(D)J Recombination. *Cell* **166**, 102-114
35. Carmona, L. M., and Schatz, D. G. (2017) New insights into the evolutionary origins of the recombination-activating gene proteins and V(D)J recombination. *Febs j* **284**, 1590-1605
36. Kirch, S. A., Sudarsanam, P., and Oettinger, M. A. (1996) Regions of RAG1 protein critical for V(D)J recombination. *Eur J Immunol* **26**, 886-891
37. Cuomo, C. A., and Oettinger, M. A. (1994) Analysis of regions of RAG-2 important for V(D)J recombination. *Nucleic Acids Res* **22**, 1810-1814
38. Sadofsky, M. J., Hesse, J. E., and Gellert, M. (1994) Definition of a core region of RAG-2 that is functional in V(D)J recombination. *Nucleic Acids Res* **22**, 1805-1809
39. Fugmann, S. D., Villey, I. J., Ptaszek, L. M., and Schatz, D. G. (2000) Identification of two catalytic residues in RAG1 that define a single active site within the RAG1/RAG2 protein complex. *Mol Cell* **5**, 97-107
40. Kim, D. R., Dai, Y., Mundy, C. L., Yang, W., and Oettinger, M. A. (1999) Mutations of acidic residues in RAG1 define the active site of the V(D)J recombinase. *Genes Dev* **13**, 3070-3080
41. Landree, M. A., Wibbenmeyer, J. A., and Roth, D. B. (1999) Mutational analysis of RAG1 and RAG2 identifies three catalytic amino acids in RAG1 critical for both cleavage steps of V(D)J recombination. *Genes Dev* **13**, 3059-3069
42. Roman, C. A., and Baltimore, D. (1996) Genetic evidence that the RAG1 protein directly participates in V(D)J recombination through substrate recognition. *Proc Natl Acad Sci U S A* **93**, 2333-2338
43. Difilippantonio, M. J., McMahan, C. J., Eastman, Q. M., Spanopoulou, E., and Schatz, D. G. (1996) RAG1 mediates signal sequence recognition and recruitment of RAG2 in V(D)J recombination. *Cell* **87**, 253-262
44. Sadofsky, M. J., Hesse, J. E., van Gent, D. C., and Gellert, M. (1995) RAG-1 mutations that affect the target specificity of V(D)j recombination: a possible direct role of RAG-1 in site recognition. *Genes Dev* **9**, 2193-2199
45. Fugmann, S. D., and Schatz, D. G. (2001) Identification of basic residues in RAG2 critical for DNA binding by the RAG1-RAG2 complex. *Mol Cell* **8**, 899-910
46. McMahan, C. J., Difilippantonio, M. J., Rao, N., Spanopoulou, E., and Schatz, D. G. (1997) A basic motif in the N-terminal region of RAG1 enhances V(D)J recombination activity. *Mol Cell Biol* **17**, 4544-4552

47. Cuomo, C. A., Kirch, S. A., Gyuris, J., Brent, R., and Oettinger, M. A. (1994) Rch1, a protein that specifically interacts with the RAG-1 recombination-activating protein. *Proc Natl Acad Sci U S A* **91**, 6156-6160
48. Spanopoulou, E., Cortes, P., Shih, C., Huang, C. M., Silver, D. P., Svec, P., and Baltimore, D. (1995) Localization, interaction, and RNA binding properties of the V(D)J recombination-activating proteins RAG1 and RAG2. *Immunity* **3**, 715-726
49. Brecht, R. M., Liu, C. C., Beilinson, H. A., Khitun, A., Slavoff, S. A., and Schatz, D. G. (2020) Nucleolar localization of RAG1 modulates V(D)J recombination activity. *Proc Natl Acad Sci U S A* **117**, 4300-4309
50. Jones, J. M., and Gellert, M. (2003) Autoubiquitylation of the V(D)J recombinase protein RAG1. *Proc Natl Acad Sci U S A* **100**, 15446-15451
51. Singh, S. K., and Gellert, M. (2015) Role of RAG1 autoubiquitination in V(D)J recombination. *Proc Natl Acad Sci U S A* **112**, 8579-8583
52. Kirch, S. A., Rathbun, G. A., and Oettinger, M. A. (1998) Dual role of RAG2 in V(D)J recombination: catalysis and regulation of ordered Ig gene assembly. *Embo j* **17**, 4881-4886
53. Elkin, S. K., Ivanov, D., Ewalt, M., Ferguson, C. G., Hyberts, S. G., Sun, Z. Y., Prestwich, G. D., Yuan, J., Wagner, G., Oettinger, M. A., and Gozani, O. P. (2005) A PHD finger motif in the C terminus of RAG2 modulates recombination activity. *J Biol Chem* **280**, 28701-28710
54. Ward, A., Kumari, G., Sen, R., and Desiderio, S. (2018) The RAG-2 Inhibitory Domain Gates Accessibility of the V(D)J Recombinase to Chromatin. *Mol Cell Biol* **38**
55. Couedel, C., Roman, C., Jones, A., Vezzoni, P., Villa, A., and Cortes, P. (2010) Analysis of mutations from SCID and Omenn syndrome patients reveals the central role of the Rag2 PHD domain in regulating V(D)J recombination. *J Clin Invest* **120**, 1337-1344
56. Papaemmanuil, E., Rapado, I., Li, Y., Potter, N. E., Wedge, D. C., Tubio, J., Alexandrov, L. B., Van Loo, P., Cooke, S. L., Marshall, J., Martincorena, I., Hinton, J., Gundem, G., van Delft, F. W., Nik-Zainal, S., Jones, D. R., Ramakrishna, M., Titley, I., Stebbings, L., Leroy, C., Menzies, A., Gamble, J., Robinson, B., Mudie, L., Raine, K., O'Meara, S., Teague, J. W., Butler, A. P., Cazzaniga, G., Biondi, A., Zuna, J., Kempski, H., Muschen, M., Ford, A. M., Stratton, M. R., Greaves, M., and Campbell, P. J. (2014) RAG-mediated recombination is the predominant driver of oncogenic rearrangement in ETV6-RUNX1 acute lymphoblastic leukemia. *Nat Genet* **46**, 116-125
57. Lin, W. C., and Desiderio, S. (1993) Regulation of V(D)J recombination activator protein RAG-2 by phosphorylation. *Science* **260**, 953-959
58. Lin, W. C., and Desiderio, S. (1994) Cell cycle regulation of V(D)J recombination-activating protein RAG-2. *Proc Natl Acad Sci U S A* **91**, 2733-2737
59. Lee, J., and Desiderio, S. (1999) Cyclin A/CDK2 regulates V(D)J recombination by coordinating RAG-2 accumulation and DNA repair. *Immunity* **11**, 771-781
60. Jiang, H., Chang, F. C., Ross, A. E., Lee, J., Nakayama, K., Nakayama, K., and Desiderio, S. (2005) Ubiquitylation of RAG-2 by Skp2-SCF links destruction of the V(D)J recombinase to the cell cycle. *Molecular cell* **18**, 699-709
61. Li, Z., Dordai, D. I., Lee, J., and Desiderio, S. (1996) A conserved degradation signal regulates RAG-2 accumulation during cell division and links V(D)J recombination to the cell cycle. *Immunity* **5**, 575-589

62. Zhang, L., Reynolds, T. L., Shan, X., and Desiderio, S. (2011) Coupling of V(D)J recombination to the cell cycle suppresses genomic instability and lymphoid tumorigenesis. *Immunity* **34**, 163-174
63. Ross, A. E., Vuica, M., and Desiderio, S. (2003) Overlapping signals for protein degradation and nuclear localization define a role for intrinsic RAG-2 nuclear uptake in dividing cells. *Molecular and cellular biology* **23**, 5308-5319
64. Steen, S. B., Han, J. O., Mundy, C., Oettinger, M. A., and Roth, D. B. (1999) Roles of the "dispensable" portions of RAG-1 and RAG-2 in V(D)J recombination. *Mol Cell Biol* **19**, 3010-3017
65. Akamatsu, Y., Monroe, R., Dudley, D. D., Elkin, S. K., Gartner, F., Talukder, S. R., Takahama, Y., Alt, F. W., Bassing, C. H., and Oettinger, M. A. (2003) Deletion of the RAG2 C terminus leads to impaired lymphoid development in mice. *Proc Natl Acad Sci U S A* **100**, 1209-1214
66. Liang, H. E., Hsu, L. Y., Cado, D., Cowell, L. G., Kelsoe, G., and Schlissel, M. S. (2002) The "dispensable" portion of RAG2 is necessary for efficient V-to-DJ rearrangement during B and T cell development. *Immunity* **17**, 639-651
67. Curry, J. D., and Schlissel, M. S. (2008) RAG2's non-core domain contributes to the ordered regulation of V(D)J recombination. *Nucleic Acids Res* **36**, 5750-5762
68. Hsu, L. Y., Lauring, J., Liang, H. E., Greenbaum, S., Cado, D., Zhuang, Y., and Schlissel, M. S. (2003) A conserved transcriptional enhancer regulates RAG gene expression in developing B cells. *Immunity* **19**, 105-117
69. Talukder, S. R., Dudley, D. D., Alt, F. W., Takahama, Y., and Akamatsu, Y. (2004) Increased frequency of aberrant V(D)J recombination products in core RAG-expressing mice. *Nucleic Acids Res* **32**, 4539-4549
70. Sekiguchi, J. A., Whitlow, S., and Alt, F. W. (2001) Increased accumulation of hybrid V(D)J joins in cells expressing truncated versus full-length RAGs. *Mol Cell* **8**, 1383-1390
71. Janeway CA Jr, T. P., Walport M, et al. (2001) The generation of diversity in immunoglobulins. in *Immunobiology: The Immune System in Health and Disease*, Garland Science, New York. pp
72. Akamatsu, Y., and Oettinger, M. A. (1998) Distinct roles of RAG1 and RAG2 in binding the V(D)J recombination signal sequences. *Mol Cell Biol* **18**, 4670-4678
73. McBlane, J. F., van Gent, D. C., Ramsden, D. A., Romeo, C., Cuomo, C. A., Gellert, M., and Oettinger, M. A. (1995) Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. *Cell* **83**, 387-395
74. Eastman, Q. M., and Schatz, D. G. (1997) Nicking is asynchronous and stimulated by synapsis in 12/23 rule-regulated V(D)J cleavage. *Nucleic Acids Res* **25**, 4370-4378
75. Ramsden, D. A., and Gellert, M. (1995) Formation and resolution of double-strand break intermediates in V(D)J rearrangement. *Genes Dev* **9**, 2409-2420
76. van Gent, D. C., Ramsden, D. A., and Gellert, M. (1996) The RAG1 and RAG2 proteins establish the 12/23 rule in V(D)J recombination. *Cell* **85**, 107-113
77. Hiom, K., and Gellert, M. (1998) Assembly of a 12/23 paired signal complex: a critical control point in V(D)J recombination. *Mol Cell* **1**, 1011-1019
78. Lee, A. I., Fugmann, S. D., Cowell, L. G., Ptaszek, L. M., Kelsoe, G., and Schatz, D. G. (2003) A functional analysis of the spacer of V(D)J recombination signal sequences. *PLoS Biol* **1**, E1

79. Shetty, K., and Schatz, D. G. (2015) Recruitment of RAG1 and RAG2 to Chromatinized DNA during V(D)J Recombination. *Mol Cell Biol* **35**, 3701-3713
80. Steen, S. B., Gomelsky, L., Speidel, S. L., and Roth, D. B. (1997) Initiation of V(D)J recombination in vivo: role of recombination signal sequences in formation of single and paired double-strand breaks. *Embo j* **16**, 2656-2664
81. Ramsden, D. A., McBlane, J. F., van Gent, D. C., and Gellert, M. (1996) Distinct DNA sequence and structure requirements for the two steps of V(D)J recombination signal cleavage. *Embo j* **15**, 3197-3206
82. Li, W., Swanson, P., and Desiderio, S. (1997) RAG-1 and RAG-2-dependent assembly of functional complexes with V(D)J recombination substrates in solution. *Mol Cell Biol* **17**, 6932-6939
83. Hesse, J. E., Lieber, M. R., Mizuuchi, K., and Gellert, M. (1989) V(D)J recombination: a functional definition of the joining signals. *Genes Dev* **3**, 1053-1061
84. Shockett, P. E., and Schatz, D. G. (1999) DNA hairpin opening mediated by the RAG1 and RAG2 proteins. *Mol Cell Biol* **19**, 4159-4166
85. Deriano, L., and Roth, D. B. (2013) Modernizing the nonhomologous end-joining repertoire: alternative and classical NHEJ share the stage. *Annu Rev Genet* **47**, 433-455
86. Bogue, M. A., Wang, C., Zhu, C., and Roth, D. B. (1997) V(D)J recombination in Ku86-deficient mice: distinct effects on coding, signal, and hybrid joint formation. *Immunity* **7**, 37-47
87. Neal, J. A., and Meek, K. (2011) Choosing the right path: does DNA-PK help make the decision? *Mutat Res* **711**, 73-86
88. Purugganan, M. M., Shah, S., Kearney, J. F., and Roth, D. B. (2001) Ku80 is required for addition of N nucleotides to V(D)J recombination junctions by terminal deoxynucleotidyl transferase. *Nucleic Acids Res* **29**, 1638-1646
89. Sandor, Z., Calicchio, M. L., Sargent, R. G., Roth, D. B., and Wilson, J. H. (2004) Distinct requirements for Ku in N nucleotide addition at V(D)J- and non-V(D)J-generated double-strand breaks. *Nucleic Acids Res* **32**, 1866-1873
90. Repasky, J. A., Corbett, E., Boboila, C., and Schatz, D. G. (2004) Mutational analysis of terminal deoxynucleotidyltransferase-mediated N-nucleotide addition in V(D)J recombination. *J Immunol* **172**, 5478-5488
91. Desiderio, S. V., Yancopoulos, G. D., Paskind, M., Thomas, E., Boss, M. A., Landau, N., Alt, F. W., and Baltimore, D. (1984) Insertion of N regions into heavy-chain genes is correlated with expression of terminal deoxytransferase in B cells. *Nature* **311**, 752-755
92. Lieber, M. R., Hesse, J. E., Mizuuchi, K., and Gellert, M. (1988) Lymphoid V(D)J recombination: nucleotide insertion at signal joints as well as coding joints. *Proc Natl Acad Sci U S A* **85**, 8588-8592
93. Lu, H., Schwarz, K., and Lieber, M. R. (2007) Extent to which hairpin opening by the Artemis:DNA-PKcs complex can contribute to junctional diversity in V(D)J recombination. *Nucleic Acids Res* **35**, 6917-6923
94. Leu, T. M., and Schatz, D. G. (1995) rag-1 and rag-2 are components of a high-molecular-weight complex, and association of rag-2 with this complex is rag-1 dependent. *Mol Cell Biol* **15**, 5657-5670
95. McMahan, C. J., Sadofsky, M. J., and Schatz, D. G. (1997) Definition of a large region of RAG1 that is important for coimmunoprecipitation of RAG2. *J Immunol* **158**, 2202-2210

96. Rodgers, K. K., Villey, I. J., Ptaszek, L., Corbett, E., Schatz, D. G., and Coleman, J. E. (1999) A dimer of the lymphoid protein RAG1 recognizes the recombination signal sequence and the complex stably incorporates the high mobility group protein HMG2. *Nucleic Acids Res* **27**, 2938-2946
97. van Gent, D. C., Hiom, K., Paull, T. T., and Gellert, M. (1997) Stimulation of V(D)J cleavage by high mobility group proteins. *Embo j* **16**, 2665-2670
98. Little, A. J., Corbett, E., Ortega, F., and Schatz, D. G. (2013) Cooperative recruitment of HMGB1 during V(D)J recombination through interactions with RAG1 and DNA. *Nucleic Acids Res* **41**, 3289-3301
99. Yin, F. F., Bailey, S., Innis, C. A., Ciubotaru, M., Kamtekar, S., Steitz, T. A., and Schatz, D. G. (2009) Structure of the RAG1 nonamer binding domain with DNA reveals a dimer that mediates DNA synapsis. *Nature structural & molecular biology* **16**, 499-508
100. Ciubotaru, M., Trexler, A. J., Spiridon, L. N., Surleac, M. D., Rhoades, E., Petrescu, A. J., and Schatz, D. G. (2013) RAG and HMGB1 create a large bend in the 23RSS in the V(D)J recombination synaptic complexes. *Nucleic Acids Res* **41**, 2437-2454
101. Swanson, P. C., and Desiderio, S. (1998) V(D)J recombination signal recognition: distinct, overlapping DNA-protein contacts in complexes containing RAG1 with and without RAG2. *Immunity* **9**, 115-125
102. Swanson, P. C., and Desiderio, S. (1999) RAG-2 promotes heptamer occupancy by RAG-1 in the assembly of a V(D)J initiation complex. *Mol Cell Biol* **19**, 3674-3683
103. Kim, M. S., Chuenchor, W., Chen, X., Cui, Y., Zhang, X., Zhou, Z. H., Gellert, M., and Yang, W. (2018) Cracking the DNA Code for V(D)J Recombination. *Mol Cell* **70**, 358-370.e354
104. Ru, H., Mi, W., Zhang, P., Alt, F. W., Schatz, D. G., Liao, M., and Wu, H. (2018) DNA melting initiates the RAG catalytic pathway. *Nat Struct Mol Biol* **25**, 732-742
105. Chen, X., Cui, Y., Best, R. B., Wang, H., Zhou, Z. H., Yang, W., and Gellert, M. (2020) Cutting antiparallel DNA strands in a single active site. *Nat Struct Mol Biol* **27**, 119-126
106. Agrawal, A., and Schatz, D. G. (1997) RAG1 and RAG2 form a stable postcleavage synaptic complex with DNA containing signal ends in V(D)J recombination. *Cell* **89**, 43-53
107. Giblin, W., Chatterji, M., Westfield, G., Masud, T., Theisen, B., Cheng, H. L., DeVido, J., Alt, F. W., Ferguson, D. O., Schatz, D. G., and Sekiguchi, J. (2009) Leaky severe combined immunodeficiency and aberrant DNA rearrangements due to a hypomorphic RAG1 mutation. *Blood* **113**, 2965-2975
108. Lieber, M. R., Hesse, J. E., Mizuuchi, K., and Gellert, M. (1987) Developmental stage specificity of the lymphoid V(D)J recombination activity. *Genes Dev* **1**, 751-761
109. Turka, L. A., Schatz, D. G., Oettinger, M. A., Chun, J. J., Gorka, C., Lee, K., McCormack, W. T., and Thompson, C. B. (1991) Thymocyte expression of RAG-1 and RAG-2: termination by T cell receptor cross-linking. *Science* **253**, 778-781
110. Lauring, J., and Schlissel, M. S. (1999) Distinct factors regulate the murine RAG-2 promoter in B- and T-cell lines. *Mol Cell Biol* **19**, 2601-2612
111. Wang, Q. F., Lauring, J., and Schlissel, M. S. (2000) c-Myb binds to a sequence in the proximal region of the RAG-2 promoter and is essential for promoter activity in T-lineage cells. *Mol Cell Biol* **20**, 9203-9211
112. Timblin, G. A., Xie, L., Tjian, R., and Schlissel, M. S. (2017) Dual Mechanism of Rag Gene Repression by c-Myb during Pre-B Cell Proliferation. *Mol Cell Biol* **37**

113. Alt, F. W., Yancopoulos, G. D., Blackwell, T. K., Wood, C., Thomas, E., Boss, M., Coffman, R., Rosenberg, N., Tonegawa, S., and Baltimore, D. (1984) Ordered rearrangement of immunoglobulin heavy chain variable region segments. *The EMBO journal* **3**, 1209-1219
114. Grawunder, U., Leu, T. M., Schatz, D. G., Werner, A., Rolink, A. G., Melchers, F., and Winkler, T. H. (1995) Down-regulation of RAG1 and RAG2 gene expression in preB cells after functional immunoglobulin heavy chain rearrangement. *Immunity* **3**, 601-608
115. Kwon, J., Morshead, K. B., Guyon, J. R., Kingston, R. E., and Oettinger, M. A. (2000) Histone acetylation and hSWI/SNF remodeling act in concert to stimulate V(D)J cleavage of nucleosomal DNA. *Mol Cell* **6**, 1037-1048
116. Lion, M., Muhire, B., Namiki, Y., Tolstorukov, M. Y., and Oettinger, M. A. (2020) Alterations in chromatin at antigen receptor loci define lineage progression during B lymphopoiesis. *Proc Natl Acad Sci U S A* **117**, 5453-5462
117. Ji, Y., Resch, W., Corbett, E., Yamane, A., Casellas, R., and Schatz, D. G. (2010) The in vivo pattern of binding of RAG1 and RAG2 to antigen receptor loci. *Cell* **141**, 419-431
118. Chen, L., Zhao, L., Alt, F. W., and Krangel, M. S. (2016) An Ectopic CTCF Binding Element Inhibits Tcrd Rearrangement by Limiting Contact between Vδ and Dδ Gene Segments. *J Immunol* **197**, 3188-3197
119. Yancopoulos, G. D., Blackwell, T. K., Suh, H., Hood, L., and Alt, F. W. (1986) Introduced T cell receptor variable region gene segments recombine in pre-B cells: evidence that B and T cells use a common recombinase. *Cell* **44**, 251-259
120. Ferrier, P., Covey, L. R., Suh, H., Winoto, A., Hood, L., and Alt, F. W. (1989) T cell receptor DJ but not VDJ rearrangement within a recombination substrate introduced into a pre-B cell line. *Int Immunol* **1**, 66-74
121. Cobb, R. M., Oestreich, K. J., Osipovich, O. A., and Oltz, E. M. (2006) Accessibility control of V(D)J recombination. *Adv Immunol* **91**, 45-109
122. Sakai, E., Bottaro, A., Davidson, L., Sleckman, B. P., and Alt, F. W. (1999) Recombination and transcription of the endogenous Ig heavy chain locus is effected by the Ig heavy chain intronic enhancer core region in the absence of the matrix attachment regions. *Proc Natl Acad Sci U S A* **96**, 1526-1531
123. Serwe, M., and Sablitzky, F. (1993) V(D)J recombination in B cells is impaired but not blocked by targeted deletion of the immunoglobulin heavy chain intron enhancer. *Embo j* **12**, 2321-2327
124. Chen, J., Young, F., Bottaro, A., Stewart, V., Smith, R. K., and Alt, F. W. (1993) Mutations of the intronic IgH enhancer and its flanking sequences differentially affect accessibility of the JH locus. *Embo j* **12**, 4635-4645
125. Bories, J. C., Demengeot, J., Davidson, L., and Alt, F. W. (1996) Gene-targeted deletion and replacement mutations of the T-cell receptor beta-chain enhancer: the role of enhancer elements in controlling V(D)J recombination accessibility. *Proc Natl Acad Sci U S A* **93**, 7871-7876
126. Sleckman, B. P., Bardon, C. G., Ferrini, R., Davidson, L., and Alt, F. W. (1997) Function of the TCR alpha enhancer in alphabeta and gammadelta T cells. *Immunity* **7**, 505-515
127. Bouvier, G., Watrin, F., Naspetti, M., Verthuy, C., Naquet, P., and Ferrier, P. (1996) Deletion of the mouse T-cell receptor beta gene enhancer blocks alphabeta T-cell development. *Proc Natl Acad Sci U S A* **93**, 7877-7881

128. Afshar, R., Pierce, S., Bolland, D. J., Corcoran, A., and Oltz, E. M. (2006) Regulation of IgH gene assembly: role of the intronic enhancer and 5'DQ52 region in targeting DHJH recombination. *J Immunol* **176**, 2439-2447
129. Perlot, T., Alt, F. W., Bassing, C. H., Suh, H., and Pinaud, E. (2005) Elucidation of IgH intronic enhancer functions via germ-line deletion. *Proc Natl Acad Sci U S A* **102**, 14362-14367
130. Inlay, M., Alt, F. W., Baltimore, D., and Xu, Y. (2002) Essential roles of the kappa light chain intronic enhancer and 3' enhancer in kappa rearrangement and demethylation. *Nat Immunol* **3**, 463-468
131. Xu, Y., Davidson, L., Alt, F. W., and Baltimore, D. (1996) Deletion of the Ig kappa light chain intronic enhancer/matrix attachment region impairs but does not abolish V kappa J kappa rearrangement. *Immunity* **4**, 377-385
132. Oestreich, K. J., Cobb, R. M., Pierce, S., Chen, J., Ferrier, P., and Oltz, E. M. (2006) Regulation of TCRbeta gene assembly by a promoter/enhancer holocomplex. *Immunity* **24**, 381-391
133. Whitehurst, C. E., Chattopadhyay, S., and Chen, J. (1999) Control of V(D)J recombinational accessibility of the D beta 1 gene segment at the TCR beta locus by a germline promoter. *Immunity* **10**, 313-322
134. Villey, I., Caillol, D., Selz, F., Ferrier, P., and de Villartay, J. P. (1996) Defect in rearrangement of the most 5' TCR-J alpha following targeted deletion of T early alpha (TEA): implications for TCR alpha locus accessibility. *Immunity* **5**, 331-342
135. Ji, Y., Little, A. J., Banerjee, J. K., Hao, B., Oltz, E. M., Krangel, M. S., and Schatz, D. G. (2010) Promoters, enhancers, and transcription target RAG1 binding during V(D)J recombination. *The Journal of experimental medicine* **207**, 2809-2816
136. Stanhope-Baker, P., Hudson, K. M., Shaffer, A. L., Constantinescu, A., and Schlissel, M. S. (1996) Cell type-specific chromatin structure determines the targeting of V(D)J recombinase activity in vitro. *Cell* **85**, 887-897
137. Subrahmanyam, R., and Sen, R. (2010) RAGs' eye view of the immunoglobulin heavy chain gene locus. *Semin Immunol* **22**, 337-345
138. Cherry, S. R., and Baltimore, D. (1999) Chromatin remodeling directly activates V(D)J recombination. *Proc Natl Acad Sci U S A* **96**, 10788-10793
139. Kwon, J., Imbalzano, A. N., Matthews, A., and Oettinger, M. A. (1998) Accessibility of nucleosomal DNA to V(D)J cleavage is modulated by RSS positioning and HMG1. *Mol Cell* **2**, 829-839
140. Golding, A., Chandler, S., Ballestar, E., Wolffe, A. P., and Schlissel, M. S. (1999) Nucleosome structure completely inhibits in vitro cleavage by the V(D)J recombinase. *Embo j* **18**, 3712-3723
141. Cherry, S. R., Beard, C., Jaenisch, R., and Baltimore, D. (2000) V(D)J recombination is not activated by demethylation of the kappa locus. *Proc Natl Acad Sci U S A* **97**, 8467-8472
142. Patenge, N., Elkin, S. K., and Oettinger, M. A. (2004) ATP-dependent remodeling by SWI/SNF and ISWI proteins stimulates V(D)J cleavage of 5 S arrays. *J Biol Chem* **279**, 35360-35367
143. Pulivarthy, S. R., Lion, M., Kuzu, G., Matthews, A. G., Borowsky, M. L., Morris, J., Kingston, R. E., Dennis, J. H., Tolstorukov, M. Y., and Oettinger, M. A. (2016)

Regulated large-scale nucleosome density patterns and precise nucleosome positioning correlate with V(D)J recombination. *Proc Natl Acad Sci U S A* **113**, E6427-e6436
144. Du, H., Ishii, H., Pazin, M. J., and Sen, R. (2008) Activation of 12/23-RSS-dependent RAG cleavage by hSWI/SNF complex in the absence of transcription. *Mol Cell* **31**, 641-649
145. Osipovich, O. A., Subrahmanyam, R., Pierce, S., Sen, R., and Oltz, E. M. (2009) Cutting edge: SWI/SNF mediates antisense Igh transcription and locus-wide accessibility in B cell precursors. *J Immunol* **183**, 1509-1513
146. Tsukada, S., Sugiyama, H., Oka, Y., and Kishimoto, S. (1990) Estimation of D segment usage in initial D to JH joinings in a murine immature B cell line. Preferential usage of DFL16.1, the most 5' D segment and DQ52, the most JH-proximal D segment. *J Immunol* **144**, 4053-4059
147. Teng, G., Maman, Y., Resch, W., Kim, M., Yamane, A., Qian, J., Kieffer-Kwon, K. R., Mandal, M., Ji, Y., Meffre, E., Clark, M. R., Cowell, L. G., Casellas, R., and Schatz, D. G. (2015) RAG Represents a Widespread Threat to the Lymphocyte Genome. *Cell* **162**, 751-765
148. Maman, Y., Teng, G., Seth, R., Kleinstein, S. H., and Schatz, D. G. (2016) RAG1 targeting in the genome is dominated by chromatin interactions mediated by the non-core regions of RAG1 and RAG2. *Nucleic Acids Res* **44**, 9624-9637
149. Brinkmann, H., Venkatesh, B., Brenner, S., and Meyer, A. (2004) Nuclear protein-coding genes support lungfish and not the coelacanth as the closest living relatives of land vertebrates. *Proc Natl Acad Sci U S A* **101**, 4900-4905
150. Grundy, G. J., Yang, W., and Gellert, M. (2010) Autoinhibition of DNA cleavage mediated by RAG1 and RAG2 is overcome by an epigenetic signal in V(D)J recombination. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 22487-22492
151. Hesse, J. E., Lieber, M. R., Gellert, M., and Mizuuchi, K. (1987) Extrachromosomal DNA substrates in pre-B cells undergo inversion or deletion at immunoglobulin V-(D)-J joining signals. *Cell* **49**, 775-783
152. Pomerantz, J. L., Denny, E. M., and Baltimore, D. (2002) CARD11 mediates factor-specific activation of NF-kappaB by the T cell receptor complex. *Embo j* **21**, 5184-5194
153. Gapud, E. J., Lee, B. S., Mahowald, G. K., Bassing, C. H., and Sleckman, B. P. (2011) Repair of chromosomal RAG-mediated DNA breaks by mutant RAG proteins lacking phosphatidylinositol 3-like kinase consensus phosphorylation sites. *J Immunol* **187**, 1826-1834
154. Zhang, Y. H., Shetty, K., Surleac, M. D., Petrescu, A. J., and Schatz, D. G. (2015) Mapping and Quantitation of the Interaction between the Recombination Activating Gene Proteins RAG1 and RAG2. *J Biol Chem* **290**, 11802-11817
155. Ru, H., Chambers, M. G., Fu, T. M., Tong, A. B., Liao, M., and Wu, H. (2015) Molecular Mechanism of V(D)J Recombination from Synaptic RAG1-RAG2 Complex Structures. *Cell* **163**, 1138-1152
156. Byrum, J. N., Zhao, S., Rahman, N. S., Gwyn, L. M., Rodgers, W., and Rodgers, K. K. (2015) An interdomain boundary in RAG1 facilitates cooperative binding to RAG2 in formation of the V(D)J recombinase complex. *Protein Sci* **24**, 861-873
157. Notarangelo, L. D., Kim, M. S., Walter, J. E., and Lee, Y. N. (2016) Human RAG mutations: biochemistry and clinical implications. *Nat Rev Immunol* **16**, 234-246

158. Villa, A., Sobacchi, C., Notarangelo, L. D., Bozzi, F., Abinun, M., Abrahamsen, T. G., Arkwright, P. D., Baniyash, M., Brooks, E. G., Conley, M. E., Cortes, P., Duse, M., Fasth, A., Filipovich, A. M., Infante, A. J., Jones, A., Mazzolari, E., Muller, S. M., Pasic, S., Rechavi, G., Sacco, M. G., Santagata, S., Schroeder, M. L., Seger, R., Strina, D., Ugazio, A., Valiaho, J., Vihinen, M., Vogler, L. B., Ochs, H., Vezzoni, P., Friedrich, W., and Schwarz, K. (2001) V(D)J recombination defects in lymphocytes due to RAG mutations: severe immunodeficiency with a spectrum of clinical presentations. *Blood* **97**, 81-88
159. Tirosh, I., Yamazaki, Y., Frugoni, F., Ververs, F. A., Allenspach, E. J., Zhang, Y., Burns, S., Al-Herz, W., Noroski, L., Walter, J. E., Gennery, A. R., van der Burg, M., Notarangelo, L. D., and Lee, Y. N. (2019) Recombination activity of human recombination-activating gene 2 (RAG2) mutations and correlation with clinical phenotype. *J Allergy Clin Immunol* **143**, 726-735
160. Hawley, R. G., Lieu, F. H., Fong, A. Z., and Hawley, T. S. (1994) Versatile retroviral vectors for potential use in gene therapy. *Gene Ther* **1**, 136-138
161. Bredemeyer, A. L., Sharma, G. G., Huang, C. Y., Helmink, B. A., Walker, L. M., Khor, K. C., Nuskey, B., Sullivan, K. E., Pandita, T. K., Bassing, C. H., and Sleckman, B. P. (2006) ATM stabilizes DNA double-strand-break complexes during V(D)J recombination. *Nature* **442**, 466-470
162. Subrahmanyam, R., and Sen, R. (2012) Epigenetic features that regulate IgH locus recombination and expression. *Curr Top Microbiol Immunol* **356**, 39-63
163. Shih, H. Y., and Krangel, M. S. (2013) Chromatin architecture, CCCTC-binding factor, and V(D)J recombination: managing long-distance relationships at antigen receptor loci. *J Immunol* **190**, 4915-4921
164. Glusman, G., Rowen, L., Lee, I., Boysen, C., Roach, J. C., Smit, A. F., Wang, K., Koop, B. F., and Hood, L. (2001) Comparative genomics of the human and mouse T cell receptor loci. *Immunity* **15**, 337-349
165. Ehlich, A., Schaal, S., Gu, H., Kitamura, D., Müller, W., and Rajewsky, K. (1993) Immunoglobulin heavy and light chain genes rearrange independently at early stages of B cell development. *Cell* **72**, 695-704
166. Winkler, T. H., and Mårtensson, I. L. (2018) The Role of the Pre-B Cell Receptor in B Cell Development, Repertoire Selection, and Tolerance. *Front Immunol* **9**, 2423
167. Nemazee, D. A., and Bürki, K. (1989) Clonal deletion of B lymphocytes in a transgenic mouse bearing anti-MHC class I antibody genes. *Nature* **337**, 562-566
168. Hartley, S. B., Crosbie, J., Brink, R., Kantor, A. B., Basten, A., and Goodnow, C. C. (1991) Elimination from peripheral lymphoid tissues of self-reactive B lymphocytes recognizing membrane-bound antigens. *Nature* **353**, 765-769
169. Hartley, S. B., Cooke, M. P., Fulcher, D. A., Harris, A. W., Cory, S., Basten, A., and Goodnow, C. C. (1993) Elimination of self-reactive B lymphocytes proceeds in two stages: arrested development and cell death. *Cell* **72**, 325-335
170. Pelanda, R., and Torres, R. M. (2012) Central B-cell tolerance: where selection begins. *Cold Spring Harb Perspect Biol* **4**, a007146
171. Cyster, J. G., and Allen, C. D. C. (2019) B Cell Responses: Cell Interaction Dynamics and Decisions. *Cell* **177**, 524-540
172. Krangel, M. S. (2009) Mechanics of T cell receptor gene rearrangement. *Curr Opin Immunol* **21**, 133-139

173. Zhang, Y., McCord, R. P., Ho, Y. J., Lajoie, B. R., Hildebrand, D. G., Simon, A. C., Becker, M. S., Alt, F. W., and Dekker, J. (2012) Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell* **148**, 908-921
174. Roix, J. J., McQueen, P. G., Munson, P. J., Parada, L. A., and Misteli, T. (2003) Spatial proximity of translocation-prone gene loci in human lymphomas. *Nat Genet* **34**, 287-291
175. Chaumeil, J., and Skok, J. A. (2012) The role of CTCF in regulating V(D)J recombination. *Curr Opin Immunol* **24**, 153-159
176. Ciccone, D. N., Namiki, Y., Chen, C., Morshead, K. B., Wood, A. L., Johnston, C. M., Morris, J. W., Wang, Y., Sadreyev, R., Corcoran, A. E., Matthews, A. G. W., and Oettinger, M. A. (2019) The murine IgH locus contains a distinct DNA sequence motif for the chromatin regulatory factor CTCF. *J Biol Chem* **294**, 13580-13592
177. Hu, J., Zhang, Y., Zhao, L., Frock, R. L., Du, Z., Meyers, R. M., Meng, F. L., Schatz, D. G., and Alt, F. W. (2015) Chromosomal Loop Domains Direct the Recombination of Antigen Receptor Genes. *Cell* **163**, 947-959
178. Jain, S., Ba, Z., Zhang, Y., Dai, H. Q., and Alt, F. W. (2018) CTCF-Binding Elements Mediate Accessibility of RAG Substrates During Chromatin Scanning. *Cell* **174**, 102-116.e114
179. Guo, C., Yoon, H. S., Franklin, A., Jain, S., Ebert, A., Cheng, H. L., Hansen, E., Despo, O., Bossen, C., Vettermann, C., Bates, J. G., Richards, N., Myers, D., Patel, H., Gallagher, M., Schlissel, M. S., Murre, C., Busslinger, M., Giallourakis, C. C., and Alt, F. W. (2011) CTCF-binding elements mediate control of V(D)J recombination. *Nature* **477**, 424-430
180. Lin, S. G., Guo, C., Su, A., Zhang, Y., and Alt, F. W. (2015) CTCF-binding elements 1 and 2 in the Igh intergenic control region cooperatively regulate V(D)J recombination. *Proc Natl Acad Sci U S A* **112**, 1815-1820
181. Chaumeil, J., Micsinai, M., Ntziachristos, P., Deriano, L., Wang, J. M., Ji, Y., Nora, E. P., Rodesch, M. J., Jeddeloh, J. A., Aifantis, I., Kluger, Y., Schatz, D. G., and Skok, J. A. (2013) Higher-order looping and nuclear organization of Tcra facilitate targeted rag cleavage and regulated rearrangement in recombination centers. *Cell Rep* **3**, 359-370
182. Zhao, L., Frock, R. L., Du, Z., Hu, J., Chen, L., Krangel, M. S., and Alt, F. W. (2016) Orientation-specific RAG activity in chromosomal loop domains contributes to Tcrd V(D)J recombination during T cell development. *J Exp Med* **213**, 1921-1936
183. Gostissa, M., Alt, F. W., and Chiarle, R. (2011) Mechanisms that promote and suppress chromosomal translocations in lymphocytes. *Annu Rev Immunol* **29**, 319-350
184. Hakim, O., Resch, W., Yamane, A., Klein, I., Kieffer-Kwon, K. R., Jankovic, M., Oliveira, T., Bothmer, A., Voss, T. C., Ansarah-Sobrinho, C., Mathe, E., Liang, G., Cobell, J., Nakahashi, H., Robbiani, D. F., Nussenzweig, A., Hager, G. L., Nussenzweig, M. C., and Casellas, R. (2012) DNA damage defines sites of recurrent chromosomal translocations in B lymphocytes. *Nature* **484**, 69-74
185. Chiarle, R., Zhang, Y., Frock, R. L., Lewis, S. M., Molinie, B., Ho, Y. J., Myers, D. R., Choi, V. W., Compagno, M., Malkin, D. J., Neuberg, D., Monti, S., Giallourakis, C. C., Gostissa, M., and Alt, F. W. (2011) Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. *Cell* **147**, 107-119
186. Hiom, K., Melek, M., and Gellert, M. (1998) DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. *Cell* **94**, 463-470
187. Vanura, K., Montpellier, B., Le, T., Spicuglia, S., Navarro, J. M., Cabaud, O., Roulland, S., Vachez, E., Prinz, I., Ferrier, P., Marculescu, R., Jäger, U., and Nadel, B. (2007) In

vivo reinsertion of excised episomes by the V(D)J recombinase: a potential threat to genomic stability. *PLoS Biol* **5**, e43
188. Tevelev, A., and Schatz, D. G. (2000) Intermolecular V(D)J recombination. *J Biol Chem* **275**, 8341-8348
189. Livak, F., and Schatz, D. G. (1996) T-cell receptor alpha locus V(D)J recombination by-products are abundant in thymocytes and mature T cells. *Mol Cell Biol* **16**, 609-618
190. Tsai, C. L., Chatterji, M., and Schatz, D. G. (2003) DNA mismatches and GC-rich motifs target transposition by the RAG1/RAG2 transposase. *Nucleic Acids Res* **31**, 6180-6190
191. Chatterji, M., Tsai, C. L., and Schatz, D. G. (2006) Mobilization of RAG-generated signal ends by transposition and insertion in vivo. *Mol Cell Biol* **26**, 1558-1568
192. Marculescu, R., Vanura, K., Montpellier, B., Roulland, S., Le, T., Navarro, J. M., Jäger, U., McBlane, F., and Nadel, B. (2006) Recombinase, chromosomal translocations and lymphoid neoplasia: targeting mistakes and repair failures. *DNA Repair (Amst)* **5**, 1246-1258
193. Curry, J. D., Schulz, D., Guidos, C. J., Danska, J. S., Nutter, L., Nussenzweig, A., and Schlissel, M. S. (2007) Chromosomal reinsertion of broken RSS ends during T cell development. *J Exp Med* **204**, 2293-2303
194. Lewis, S. M., Agard, E., Suh, S., and Czyzyk, L. (1997) Cryptic signals and the fidelity of V(D)J joining. *Mol Cell Biol* **17**, 3125-3136
195. Marculescu, R., Le, T., Simon, P., Jaeger, U., and Nadel, B. (2002) V(D)J-mediated translocations in lymphoid neoplasms: a functional assessment of genomic instability by cryptic sites. *J Exp Med* **195**, 85-98
196. Onozawa, M., and Aplan, P. D. (2012) Illegitimate V(D)J recombination involving nonantigen receptor loci in lymphoid malignancy. *Genes Chromosomes Cancer* **51**, 525-535
197. Tsai, A. G., and Lieber, M. R. (2010) Mechanisms of chromosomal rearrangement in the human genome. *BMC Genomics* **11 Suppl 1**, S1
198. Hirokawa, S., Chure, G., Belliveau, N. M., Lovely, G. A., Anaya, M., Schatz, D. G., Baltimore, D., and Phillips, R. (2020) Sequence-dependent dynamics of synthetic and endogenous RSSs in V(D)J recombination. *Nucleic Acids Res*
199. Mijušković, M., Chou, Y. F., Gigi, V., Lindsay, C. R., Shestova, O., Lewis, S. M., and Roth, D. B. (2015) Off-Target V(D)J Recombination Drives Lymphomagenesis and Is Escalated by Loss of the Rag2 C Terminus. *Cell Rep* **12**, 1842-1852
200. Deriano, L., Chaumeil, J., Coussens, M., Multani, A., Chou, Y., Alekseyenko, A. V., Chang, S., Skok, J. A., and Roth, D. B. (2011) The RAG2 C terminus suppresses genomic instability and lymphomagenesis. *Nature* **471**, 119-123
201. Chaumeil, J., Micsinai, M., Ntziachristos, P., Roth, D. B., Aifantis, I., Kluger, Y., Deriano, L., and Skok, J. A. (2013) The RAG2 C-terminus and ATM protect genome integrity by controlling antigen receptor gene cleavage. *Nat Commun* **4**, 2231
202. Zhang, M., and Swanson, P. C. (2008) V(D)J recombinase binding and cleavage of cryptic recombination signal sequences identified from lymphoid malignancies. *J Biol Chem* **283**, 6717-6727
203. Arnal, S. M., Holub, A. J., Salus, S. S., and Roth, D. B. (2010) Non-consensus heptamer sequences destabilize the RAG post-cleavage complex, making ends available to alternative DNA repair pathways. *Nucleic Acids Res* **38**, 2944-2954

204. Rooney, S., Alt, F. W., Lombard, D., Whitlow, S., Eckersdorff, M., Fleming, J., Fugmann, S., Ferguson, D. O., Schatz, D. G., and Sekiguchi, J. (2003) Defective DNA repair and increased genomic instability in Artemis-deficient murine cells. *J Exp Med* **197**, 553-565
205. Lee, G. S., Neiditch, M. B., Salus, S. S., and Roth, D. B. (2004) RAG proteins shepherd double-strand breaks to a specific pathway, suppressing error-prone repair, but RAG nicking initiates homologous recombination. *Cell* **117**, 171-184
206. Gigi, V., Lewis, S., Shestova, O., Mijušković, M., Deriano, L., Meng, W., Luning Prak, E. T., and Roth, D. B. (2014) RAG2 mutants alter DSB repair pathway choice in vivo and illuminate the nature of 'alternative NHEJ'. *Nucleic Acids Res* **42**, 6352-6364
207. Corneo, B., Wendland, R. L., Deriano, L., Cui, X., Klein, I. A., Wong, S. Y., Arnal, S., Holub, A. J., Weller, G. R., Pancake, B. A., Shah, S., Brandt, V. L., Meek, K., and Roth, D. B. (2007) Rag mutations reveal robust alternative end joining. *Nature* **449**, 483-486
208. Fisher, M. R., Rivera-Reyes, A., Bloch, N. B., Schatz, D. G., and Bassing, C. H. (2017) Immature Lymphocytes Inhibit Rag1 and Rag2 Transcription and V(D)J Recombination in Response to DNA Double-Strand Breaks. *J Immunol* **198**, 2943-2956
209. Rodgers, W., Byrum, J. N., Sapkota, H., Rahman, N. S., Cail, R. C., Zhao, S., Schatz, D. G., and Rodgers, K. K. (2015) Spatio-temporal regulation of RAG2 following genotoxic stress. *DNA Repair (Amst)* **27**, 19-27
210. Panchakshari, R. A., Zhang, X., Kumar, V., Du, Z., Wei, P. C., Kao, J., Dong, J., and Alt, F. W. (2018) DNA double-strand break response factors influence end-joining features of IgH class switch and general translocation junctions. *Proc Natl Acad Sci U S A* **115**, 762-767
211. Unniraman, S., and Schatz, D. G. (2006) AID and Igh switch region-Myc chromosomal translocations. *DNA Repair (Amst)* **5**, 1259-1264
212. Hu, J., Tepsuporn, S., Meyers, R. M., Gostissa, M., and Alt, F. W. (2014) Developmental propagation of V(D)J recombination-associated DNA breaks and translocations in mature B cells via dicentric chromosomes. *Proc Natl Acad Sci U S A* **111**, 10269-10274
213. Gostissa, M., Ranganath, S., Bianco, J. M., and Alt, F. W. (2009) Chromosomal location targets different MYC family gene members for oncogenic translocations. *Proc Natl Acad Sci U S A* **106**, 2265-2270
214. Gostissa, M., Yan, C. T., Bianco, J. M., Cogné, M., Pinaud, E., and Alt, F. W. (2009) Long-range oncogenic activation of Igh-c-myc translocations by the Igh 3' regulatory region. *Nature* **462**, 803-807
215. Mijušković, M., Brown, S. M., Tang, Z., Lindsay, C. R., Efstathiadis, E., Deriano, L., and Roth, D. B. (2012) A streamlined method for detecting structural variants in cancer genomes by short read paired-end sequencing. *PLoS One* **7**, e48314
216. Hu, J., Meyers, R. M., Dong, J., Panchakshari, R. A., Alt, F. W., and Frock, R. L. (2016) Detecting DNA double-stranded breaks in mammalian genomes by linear amplification-mediated high-throughput genome-wide translocation sequencing. *Nat Protoc* **11**, 853-871
217. Frock, R. L., Hu, J., Meyers, R. M., Ho, Y. J., Kii, E., and Alt, F. W. (2015) Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat Biotechnol* **33**, 179-186
218. Crosetto, N., Mitra, A., Silva, M. J., Bienko, M., Dojer, N., Wang, Q., Karaca, E., Chiarle, R., Skrzypczak, M., Ginalski, K., Pasero, P., Rowicka, M., and Dikic, I. (2013)

Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nat Methods* **10**, 361-365
219. Canela, A., Sridharan, S., Sciascia, N., Tubbs, A., Meltzer, P., Sleckman, B. P., and Nussenzweig, A. (2016) DNA Breaks and End Resection Measured Genome-wide by End Sequencing. *Mol Cell* **63**, 898-911
220. Dorsett, Y., Zhou, Y., Tubbs, A. T., Chen, B. R., Purman, C., Lee, B. S., George, R., Bredemeyer, A. L., Zhao, J. Y., Sodergen, E., Weinstock, G. M., Han, N. D., Reyes, A., Oltz, E. M., Dorsett, D., Misulovin, Z., Payton, J. E., and Sleckman, B. P. (2014) HCoDES reveals chromosomal DNA end structures with single-nucleotide resolution. *Mol Cell* **56**, 808-818
221. Halper-Stromberg, E., Steranka, J., Giraldo-Castillo, N., Fuller, T., Desiderio, S., and Burns, K. H. (2013) Fine mapping of V(D)J recombinase mediated rearrangements in human lymphoid malignancies. *BMC Genomics* **14**, 565
222. Li, H., and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760
223. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079
224. Faust, G. G., and Hall, I. M. (2014) SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics* **30**, 2503-2505
225. Robinson, J. T., Thorvaldsdóttir, H., Wenger, A. M., Zehir, A., and Mesirov, J. P. (2017) Variant Review with the Integrative Genomics Viewer. *Cancer Res* **77**, e31-e34
226. Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., and Mesirov, J. P. (2011) Integrative genomics viewer. *Nat Biotechnol* **29**, 24-26
227. Thorvaldsdóttir, H., Robinson, J. T., and Mesirov, J. P. (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**, 178-192
228. (2020) Robo2 gene.
229. Layer, R. M., Chiang, C., Quinlan, A. R., and Hall, I. M. (2014) LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol* **15**, R84
230. Lieber, M. R. (2016) Mechanisms of human lymphoid chromosomal translocations. *Nat Rev Cancer* **16**, 387-398
231. Rustad, E. H., Hultcrantz, M., Yellapantula, V. D., Akhlaghi, T., Ho, C., Arcila, M. E., Roshal, M., Patel, A., Chen, D., Devlin, S. M., Jacobsen, A., Huang, Y., Miller, J. E., Papaemmanuil, E., and Landgren, O. (2019) Baseline identification of clonal V(D)J sequences for DNA-based minimal residual disease detection in multiple myeloma. *PLoS One* **14**, e0211600
232. Wang, J. H., Gostissa, M., Yan, C. T., Goff, P., Hickernell, T., Hansen, E., Difilippantonio, S., Wesemann, D. R., Zarrin, A. A., Rajewsky, K., Nussenzweig, A., and Alt, F. W. (2009) Mechanisms promoting translocations in editing and switching peripheral B cells. *Nature* **460**, 231-236
233. Küppers, R., and Dalla-Favera, R. (2001) Mechanisms of chromosomal translocations in B cell lymphomas. *Oncogene* **20**, 5580-5594
234. Li, H. (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987-2993

Curriculum Vitae

MEILING R. MAY

cell: 339-933-0352 | meilingrmay@gmail.com

RESEARCH EXPERIENCE

AstraZeneca, Graduate Student Intern **January 2020-April 2020**

Cell Culture and Fermentation Sciences Department, Biopharmaceutical Development

- Initiated investigations into molecular mechanisms underlying promising in-house media formulations for CAR-T cell cultures
- Documented and presented results to team via oral and written communications
- Transduced, expanded, and phenotyped primary T cells ex vivo for CAR-T development

Johns Hopkins University School of Medicine, Ph.D. Candidate **2014-2020**

PI: Stephen Desiderio M.D. Ph.D. and Carol Greider Ph.D.

- Initiated and led a project which determined the epigenetic regulation of V(D)J Recombination and its role in lymphomagenesis, which resulted in 2 presentations at major conferences
- Developed and optimized flow cytometry-based assay for quantifying V(D)J recombination
- Designed and optimized a capture-based NGS approach to identify novel translocations and genomic rearrangements in a mouse model of mistimed V(D)J Recombination and performed bioinformatic analyses
- Created knock-out and knock-in cell lines and mouse models utilizing CRISPR/Cas9 technology
- Secured a $500 travel grant to attend American Society for Hematology meeting

College of the Holy Cross, Undergraduate Researcher; Summer Fellow **2012-2014**

PI: Ann Sheehy Ph.D.

- Investigated the antiviral properties of human APOBEC3G and its role in an HIV-1 infection which resulted in three presentations at major conferences, 2 presentations at undergraduate research symposia, and an honors thesis
- Mentored and trained 4 junior scientists in both wet-lab practices and presentations

EDUCATION

Johns Hopkins University School of Medicine **2014- 2020**

Ph.D. candidate Biochemistry, Cellular, and Molecular Biology

Preparing Future Faculty Teaching Academy Certificate **2016**

College of the Holy Cross **2010-2014**

B.A. Biology, Summa Cum Laude

College Honors Program (only 32 of 680 students selected)

TECHNICAL SKILLS

Cell culture: CAR-T ex-vivo expansion(G-REX), mammalian cell lines, primary human and murine cultures, transfection, viral production and transduction, CRISPR/Cas9 editing

Molecular biology/biochemistry: Western blotting, Southern blotting, molecular cloning, qPCR, mutagenesis, DNA isolation/purification, RNA isolation/purification, mammalian protein purification, ELISA, in vitro biochemical assays

In vivo models: flow cytometry, FACS, CRISPR/Cas9 mouse creation, mouse colony maintenance

Bioinformatics: Illumina sequencing, NGS analysis, basic bash scripting, basic R scripting

LEADERSHIP/MANAGEMENT EXPERIENCE

Baltimore Brain Connect, Co-founder and Organizer **2018, 2019**

- Developed and lead Brain Awareness Week demonstrations and lesson plans which were implemented in over 25 elementary school classrooms in Baltimore City
- Recruited, organized, and managed 100+ volunteers to lead classroom visits which culminated in a day-long brain festival that had 400+ attendees and was featured on WYPR radio
- Secured over $500 in grants to pilot this program, created timeline and goals

Science Outside the Lines, Founding Member and Organizer **2017-2019**

- Initiated and founded a 4th grade afterschool arts integration science program in collaboration with Henderson-Hopkins elementary school and Johns Hopkins
- Secured over $3000 in grants to pilot this program, set timelines and met deadlines

Project Bridge, Co-president **2016-2018**

- Lead over 100 graduate student and post-doctoral volunteers in science outreach and science communication efforts
- Organized and executed monthly science demonstrations at local farmers market as committee head for Science at the Market efforts

MENTORING/TEACHING EXPERIENCE

Johns Hopkins University, Instructor **Winter 2018, Winter 2019**

- Designed, organized, and taught the course "How Viruses Shape Our World" to 50 undergraduate students
- Delivered active-learning styled lectures, facilitated in-class discussions and group work, and utilized novel pedagogical technology

Goucher College, Collaborative Teaching Fellow **Spring 2019**

- Developed and presented pre-laboratory lectures, created and graded assessments, and assisted instructor and students during the laboratory experiments

Johns Hopkins University School of Medicine

- Graded and lectured as teaching assistant for graduate Molecular Biology course 2017, 2018
- Developed curriculum for inducible expression systems in Genetics Graduate Course 2017
- Tutored students in graduate level Genetics course on 1:1 basis 2016-2018

PUBLICATIONS

MR May, JT Bettridge, SV Desiderio. 2020. H3K4me3 binding and relay of an allosteric signal to RAG-1 are separable functions of the RAG-2 PHD finger. J Biol Chemistry. May 2020 PMC7335790

MR May, SV Desiderio. 2020. An antigen receptor-specific capture sequencing method reveals structural variations in a mouse model of mistimed V(D)J Recombination. BMC Genomics *In Preparation- Anticipated submission Aug 2020*

ABSTRACTS AND PRESENTATIONS

MR May, JT Bettridge, SV Desiderio. 2018. Oral Abstract. An Allosteric Mechanism for Epigenetic Activation of the V(D)J Recombinase. American Society for Hematology. San Diego CA

MR May, JT Bettridge, SV Desiderio. 2018. Poster. An Allosteric Mechanism for Epigenetic Activation of the V(D)J Recombinase. American Society of Hematology Meeting on Lymphoma Biology. Washington D.C.

NK Hussain, **MR May**, KJ Monk, S Magsamen, RL Huganir. 2018. Poster. Baltimore Brain Connect: A Multi-layered Approach to Increasing Public Engagement with Neuroscience and Research. Society for Neuroscience. San Diego CA

MR May. 2014. Poster. Characterization of the Antiviral Function of APOBEC3G. Annual American Association for the Advancement of Science Conference. Chicago, IL.

MR May. 2014. Oral Abstract and invited panelist. Characterizing Novel N-terminal Domains of APOBEC3G. Conference on Retroviruses and Opportunistic Infections. Boston, MA.

*Sheehy, AM and **MR May.*** 2012. Poster. Characterizing the anti-viral function of APOBEC3G. XIX International AIDS conference. Washington D.C.

SELECT AWARDS AND HONORS

2019	**Society for Neuroscience Next Generation Award** – national award for outstanding contribution to public communication, outreach, and education about neuroscience
2019	**Baltimore Science Slam** – 1st place in science communication competition
2018	**Johns Hopkins Graduate Student Association Travel Award** – granted for American Society of Hematology meeting
2018	**Johns Hopkins Career Champion Award** – nominated for career and student support
2018	**Johns Hopkins School of Medicine Student Group Community Service Award**
2017, 2018	**Johns Hopkins Student Group Leader of the year**- finalist twice
2015	**National Science Foundation Graduate Research Fellowship Program Honorable Mention** – prestigious national research award granted to top PhD students
2011-2014	**Charles A. Dana Scholarship** – prestigious grant for top performing undergraduates in academics and service
2014	**The Rev. John W. Flavin S.J. Award in Biology** – Only award given to a graduating senior biology major for academic excellence and contributions to the biology dept.

2012 **Theresa A. Churilla Biology Book Award** – Only award given to a sophomore biology major for academic excellence and contributions to the biology dept.